1 Large Deviations and Statistical Ensembles

References:

1. A Modern Course in Statistical Physics by Reichl
2. Statistical Mechanics by K.S. Huang,
3. Entropy, Large Deviations, and Statistical Mechanics by R. Ellis.

Three-door show: emphasize the 1/3 probability is our assumption.
Homework: What if we observed 1/2 and 1/2 probability upon switching? What would be "prior" probability?

1.1 A Conditional Limit Theorem

For example, given a loaded (unfair) die, how can you determine the actual probability of each face?

In the notation introduced above: Assign $\rho_k = \frac{1}{6}$ if no additional information available. For n tosses, we obtain a configuration $\omega \in \{1, 2, \cdots, 6\}^n$

We ask the following question: Can we calculate the probabilities of 6 faces based on the additional information, say, the total sum of the values of the faces

$$S_n(\omega) \equiv \sum_{j=1}^{n} X_j(\omega) = \sum_{j=1}^{n} \omega_j?$$

The answer is yes, we can obtain asymptotic results as $n \to \infty$.

1.1.1 Setup

$\Lambda = \{y_1, \cdots, y_N\}$ — possible outcomes of random experiments

$y_1 < y_2 < \cdots < y_N, \quad y_k \in \mathbb{R}$

with probabilities: $\rho_1, \rho_2, \cdots, \rho_N$; $\rho_k > 0, \quad \sum_{k=1}^{N} \rho_k = 1$

$$\rho \equiv (\rho_1, \rho_2, \cdots, \rho_N) \in \mathcal{P}_N \equiv \left\{ \gamma \in \mathbb{R}^N | \gamma = (\gamma_1, \cdots, \gamma_N) \geq 0, \sum_{k=1}^{N} \gamma_k = 1 \right\}$$

$\forall \gamma \in \mathcal{P}_N$ defines a probability measure on the set of subsets of Λ via

$$\gamma = \gamma(dy) \equiv \sum_{k=1}^{N} \gamma_k \delta_{y_k}(dy)$$

$$y \in \Lambda,\ \delta_{y_k}(dy) = \begin{cases} 1 & \text{if } y = y_k \\ 0 & \text{otherwise} \end{cases}$$

$B \subset \Lambda$,

$$\gamma\{B\} = \sum_{y_k \in B} \gamma_k$$

For n trials, we define $\Omega_n \equiv \Lambda^n \quad \omega = (\omega_1, \cdots, \omega_n)$

$\forall \omega \in \Omega_n$ Define:

$$P_n\{\omega\} \equiv \prod_{j=1}^{n} \rho\{\omega_j\}$$

$$B \subset \Omega_n,\ P_n\{B\} \equiv \sum_{\omega \in B} P_n\{\omega\}$$

$P_n\{B\}$ is the product measure with 1-dimensional marginal ρ.

Define coordinates:

$$X_j(\omega) \equiv \omega_j, \quad j = 1, 2, \cdots, n$$

which is i.i.d. with ρ.

$\omega \in \Omega_n,\ y \in \Lambda$, Relative frequency of y in configuration ω :

$$L_n(y) \equiv L_n(\omega, y) \equiv \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j(\omega)}\{y\}$$

i.e., $\frac{1}{n} \times$ (the number of events: $\omega_j = y,\ j = 1, 2, \cdots, n$).

The sample mean of the i.i.d random variable $\left(\delta_{X_j(\omega)}\{y_1\}, \cdots, \delta_{X_j(\omega)}\{y_N\}\right)$:

$$L_n \equiv L_n(\omega) \equiv (L_n(\omega, y_1), L_n(\omega, y_2), \cdots, L_n(\omega, y_N))$$

i.e., L_n takes value in $\mathcal{P}_N$.

1.1.2 Conditional Limit Theorem

A **conditional limit theorem** for L_n.

Note that for a fair die, the sample mean

$$\frac{S_n(\omega)}{n} \simeq \text{theoretical mean } \bar{y} \equiv \sum_{k=1}^{6} k\rho_k = 3.5$$

Assume we have observed $\frac{S_n(\omega)}{n} \in [z-\delta, z]$, where δ is small, $\delta > 0$. (For the time being, we assume $1 \leq z-\delta < z < \bar{y}$, similar results hold for $\frac{S_n(\omega)}{n} \in [z, z+\delta]$, $\bar{y} \leq z < z+\delta \leq 6$.) What we need to determine is

$$\begin{aligned} \{\rho_k^*, k = 1, 2, \cdots, 6\}, \quad \sum \rho_k^* &= 1, \quad \text{such that} \\ \rho_k^* &= \lim_{n\to\infty} P_n \left\{ X_1 = k \left| \frac{S_n}{n} \in [z-\delta, z] \right. \right\} \end{aligned}$$

or determine the most probable configuration:

$$\rho^* = (\rho_1^*, \cdots, \rho_6^*)$$

of L_n, i.e., (in the setting of die, $N = 6$)

$$\begin{aligned} \rho^* &\in \mathcal{P}_N, \quad \text{such that } \forall \varepsilon > 0 \\ \lim_{n\to\infty} P_n \left\{ L_n \in B(\rho^*, \varepsilon) \left| \frac{S_n}{n} \in [z-\delta, z] \right. \right\} &= 1 \end{aligned}$$

where $B(\rho^*, \varepsilon)$ is the open ball:

$$B(\rho^*, \varepsilon) = \{\gamma \in \mathcal{P}_N \quad | \quad \|\gamma - \rho^*\| < \varepsilon\}$$

where $\|\cdots\|$ denotes the Euclidean norm in $\mathbb{R}^N$.

More generally,

$$\begin{aligned} \Lambda &= \{y_1, \cdots, y_N\} \\ &(\rho_1, \cdots, \rho_N) \in \mathcal{P}_N \end{aligned}$$

and

$$S_n \equiv \sum_{j=1}^{n} X_j, \quad \bar{y} \equiv \sum_{k=1}^{N} y_k \rho_k$$

For sufficiently small $\delta > 0$, fix an interval $[z-\delta, z] \subset [y_1, \bar{y}]$ (note that a similar theorem holds for $[z-\delta, z] \subset [\bar{y}, y_N]$

Theorem:

1. $\exists \bar{\rho} \in \mathcal{P}_N$ such that $\forall \varepsilon > 0$

$$\lim_{n\to\infty} P_n \left\{ L_n \in B\left(\bar{\rho}, \varepsilon\right) \;\middle|\; \frac{S_n}{n} \in [z-\delta, z] \right\} = 1$$

$\bar{\rho} = (\bar{\rho}_1, \cdots, \bar{\rho}_N)$ has the form

$$\bar{\rho}_k = \frac{\rho_k e^{-\beta y_k}}{\sum_{i=1}^N \rho_i e^{-\beta y_i}}$$

in which β is determined by

$$\sum_{k=1}^N y_k \bar{\rho}_k = z$$

Note that β is a function of z.

2. For any continuous function $f : \mathcal{P}_N \to \mathbb{R}$

$$\lim_{n\to\infty} \mathbf{E}_{P_n} \left\{ f\left(L_n\right) \;\middle|\; \frac{S_n}{n} \in [z-\delta, z] \right\} = f\left(\bar{\rho}\right)$$

$$\bar{\rho}_k = \lim_{n\to\infty} P_n \left\{ X_1 = y_k \;\middle|\; \frac{S_n}{n} \in [z-\delta, z] \right\}$$

Note that 2^o is the immediate consequence of 1^o with the continuity of f. Moreover, we comment that $\bar{\rho}$ is well defined. This can be seen as follows. For $\alpha \in \mathbb{R},$ define the partition function

$$Z\left(\alpha\right) \equiv \log\left(\sum_{k=1}^N \rho_k e^{\alpha y_k}\right)$$

then, we can easily verify that

$$Z'\left(-\beta\right) = \sum_{k=1}^N y_k \bar{\rho}_k$$

and that

1. $Z''\left(\alpha\right) > 0$

$$\begin{array}{rclll} Z'\left(\alpha\right) & \to & y_1 & \text{as} & \alpha \to -\infty \\ Z'\left(0\right) & = & \bar{y} & & \\ Z'\left(\alpha\right) & \to & y_N & \text{as} & \alpha \to +\infty \end{array}$$

Therefore,

1. $\exists ! \beta$ such that $Z'(-\beta) = \sum_{k=1}^{N} y_k \bar{\rho}_k = z$, where z is a constant

2. $\because y_1 < z < \bar{y}$
   $\therefore \ \beta = \beta(z) > 0$

Note that

$$\sum_{k=1}^{N} y_k \bar{\rho}_k = z = -\frac{d}{d\beta}\left(\log \sum \rho_k e^{-\beta y_k}\right)$$

1.2 Entropies

Consider N events, let us compare two situations:

1. if all the N events are equally probable;

2. one of the N events have much high probability to occur, (in the limitng case, for example, it has probability one, the rest zero)

Obviously, we would regard the first situation more uncertain. How do we quantify uncertainty? Everybody probably has heard the word entropy one way or another. Many lines of thoughts, from statistical physics in the nineteenth century to communication theory c.a. WWII, all arrived at the same mathematical expression in quantifying uncertainty. Why? Here we will present the logical simplicity and inevitabilty of it (of course, nothing is inevitable or that simple.)

1.2.1 Basic Results:

Entropy Entropy is defined as

$$S(\rho_1, \cdots, \rho_2) = -\sum_{i=1}^{N} \rho_i \log \rho_i$$

which is a measure of uncertainty. A natural question arises why we have this definition for quantifying uncertainties. The following theorem provides an answer why this is a natural characterization.

Theorem: If a function $H_N : \mathcal{P}_N \mapsto \mathbb{R}$, where

$$\mathcal{P}_N \equiv \left\{ \gamma \in \mathbb{R}^N | \gamma = (\gamma_1, \cdots, \gamma_N) \geq 0, \sum_{k=1}^{N} \gamma_k = 1 \right\}$$

satisfies the following propositions:

1. $H_N(\rho_1, \cdots, \rho_N)$ is a continuous function;
2. $A(N) \equiv H_N\left(\frac{1}{N}, \cdots, \frac{1}{N}\right)$ is monotonic increasing in N;
3. If the sample space $\Lambda = \{y_1, \cdots, y_N\}$ is divided into two subsets:

$$\Lambda_1 = \{y_1, \cdots, y_k\}, \quad \Lambda_2 = \{y_{k+1}, \cdots, y_N\}$$

with probability $Q_1 = \rho_1 + \cdots + \rho_k$, $Q_2 = \rho_{k+1} + \cdots + \rho_N$, i.e., the conditional probability is

$$\left(\frac{\rho_1}{Q_1}, \cdots \frac{\rho_k}{Q_1}\right) \quad \text{and} \quad \left(\frac{\rho_{k+1}}{Q_2}, \cdots \frac{\rho_N}{Q_2}\right)$$

then

$$\begin{aligned} H_N(\rho_1, \cdots, \rho_N) &= H_2(Q_1, Q_2) \\ &\quad + Q_1 H_k\left(\frac{\rho_1}{Q_1}, \cdots \frac{\rho_k}{Q_1}\right) \\ &\quad + Q_2 H_{N-k}\left(\frac{\rho_{k+1}}{Q_2}, \cdots \frac{\rho_N}{Q_2}\right) \end{aligned}$$

Then,

$$H_N(\rho_1, \cdots, \rho_N) = cS(\rho_1, \cdots, \rho_N) = -c \sum_{i=1}^{N} \rho_i \log \rho_i, \quad c > 0 \tag{1}$$

Note that Proposition (1) above is natural requirement and Proposition (2) reflects the following intuition: If we are confronted with two situations: (i) there are 100 equally probable events; (ii) there are one million equally probable events, which appears more uncertain to us? Shouldn't be case (ii)?

Proof: Since $H_N(\rho_1, \cdots, \rho_N)$ is continuous, it is sufficient to prove Eq. (1) holds for $q_i \in \mathbb{Q}$,. A set of any rational $q_1, \cdots q_N$ such that

$$0 \leq q_i \leq 1 \qquad \sum_{i=1}^{N} q_i = 1$$

can be written as

$$q_i = \frac{\varpi_i}{M}, \quad M = \sum_{i=1}^{N} \varpi_i$$

We consider the sample space $\Lambda = \{y_1, \cdots, y_M\}$ with the probability $\{\rho_1, \cdots, \rho_M\}$ and split Λ into N subsets:

$$\Lambda_i = \left\{y_{k_{i-1}+1}, \cdots, y_{k_i}\right\}, \quad i = 1, \cdots, N$$

with $k_0 \equiv 0$, $k_i = k_{i-1} + \varpi_i$, $\ i = 1, \cdots, N$. The probability associated with Λ_i is

$$Q_i = \rho_{k_{i-1}+1} + \cdots + \rho_{k_i}$$

From Property 3 above, we have

$$H_M\left(\rho_1, \cdots, \rho_M\right) = H_N\left(Q_1, \cdots, Q_N\right) + \sum_{i=1}^{N} Q_i H_{\varpi_i}\left(\frac{\rho_{k_{i-1}+1}}{Q_i}, \cdots \frac{\rho_{k_i}}{Q_i}\right) \tag{2}$$

If $\rho_i = \frac{1}{M}, i = 1, \cdots, M$, then

$$Q_i = \frac{\varpi_i}{M} = q_i$$

and Eq. (2) becomes

$$H_M\left(\frac{1}{M}, \cdots, \frac{1}{M}\right) = H_N\left(Q_1, \cdots, Q_N\right) + \sum_{i=1}^{N} Q_i H_{\varpi_i}\left(\frac{1}{\varpi_i}, \cdots \frac{1}{\varpi_i}\right)$$

Since $A(N) = H_N\left(\frac{1}{N}, \cdots, \frac{1}{N}\right)$, we have

$$A(M) = H_N\left(Q_1, \cdots, Q_N\right) + \sum_{i=1}^{N} Q_i A\left(\varpi_i\right) \tag{3}$$

If we set $\varpi_i = \omega$, then

$$M = N\varpi, \quad Q_i = \frac{\varpi}{M} = \frac{1}{N} \quad \forall i$$

and Eq. (3) becomes

$$A(N\varpi) = A(N) + A(\varpi)$$

Since the only continuous functional solution $A(N)$ for the above equation is

$$A(N) = c \log N$$

The constant $c > 0$ because $A(N)$ is a monotonic increasing function of N. In general, with this solution for $A(N)$, Eq. (3) gives

$$\begin{aligned} H_N\left(Q_1, \cdots, Q_N\right) &= c \log M - c \sum_{i=1}^{N} Q_i \log \varpi_i \\ &= -c \sum_{i=1}^{N} Q_i \log Q_i \end{aligned}$$

This is precisely what we desire to show.

Relative Entropy Definition of Relative Entropy

The relative entropy of $\gamma \in \mathcal{P}_N$ with respect to $\rho \in \mathcal{P}_N$ is

$$I_\rho(\gamma) \equiv \sum_{i=1}^{N} \gamma_i \log \frac{\gamma_i}{\rho_i}$$

Basic properties:

1. Non-negativity:
$$I_\rho(\gamma) \geq 0$$
The equality holds if and only if $\gamma = \rho$. i.e., $I_\rho(\gamma)$ attains its inf of 0 over $\mathcal{P}_N$ at the unique measure $\gamma = \rho$.

2. Convexity: $I_\rho(\gamma)$ is strictly convex in $\mathcal{P}_N$

Proof: 1)
$$\because \quad x \log x \geq x - 1$$
where the equality holds if and only if $x = 1$. Therefore
$$\left(\frac{\gamma_k}{\rho_k}\right) \log \frac{\gamma_k}{\rho_k} \geq \frac{\gamma_k}{\rho_k} - 1 \tag{4}$$
Multiplying ρ_k on the both sides, and then summing over k yields
$$\begin{aligned} \sum_k \gamma_k \log \frac{\gamma_k}{\rho_k} \quad &\geq \quad \sum_k \gamma_k - \sum_k \rho_k = 0 \\ \therefore \quad & \quad I_\rho(\gamma) \geq 0 \end{aligned}$$
where the equality holds if and only if $\rho = \gamma$.

2) Strict convexity of $I_\rho(\gamma)$ follows from the strict convexity of $x \log x$, $x > 0$.

Theorem: $\forall \gamma \in \mathcal{P}_N$, $\forall$ small $\varepsilon > 0$,

$$P_n\{L_n \in B(\gamma, \varepsilon)\} \approx e^{-nI_\rho(\gamma)}$$

as $n \to \infty$.

Essence of Proof:

$$\begin{aligned} P_n\{L_n \in B(\gamma, \varepsilon)\} &= P_n\left\{\omega \in \Omega_n : L_n(\omega) \sim \frac{1}{n}(n\gamma_1, \cdots n\gamma_N)\right\} \\ &\approx P_n\{\text{the number of } \{\omega_j = y_1\} \sim n\gamma_1, \cdots \text{the number of } \{\omega_j = y_N\} \sim n\gamma_N\} \\ &\approx \frac{n!}{(n\gamma_1)!\,(n\gamma_2)! \cdots (n\gamma_N)!} \rho_1^{n\gamma_1} \rho_2^{n\gamma_2} \cdots \rho_N^{n\gamma_N} \end{aligned}$$

Using Sterling's formula

$$\log n! = n\log n - n + O(\log n)$$

we have

$$\begin{aligned}
\frac{1}{n}\log P_n\{L_n \in B(\gamma,\varepsilon)\} &\approx \frac{1}{n}\log\left(\frac{n!}{(n\gamma_1)!\,(n\gamma_2)!\cdots(n\gamma_N)!}\right) + \sum_{k=1}^{N}\gamma_k\log\rho_k \\
&= -\sum_{k=1}^{N}\gamma_k\log\gamma_k + O\left(\frac{\log n}{n}\right) + \sum_{k=1}^{N}\gamma_k\log\rho_k \\
&= -\sum_{k}\gamma_k\log\frac{\gamma_k}{\rho_k} + O\left(\frac{\log n}{n}\right) \\
&= -I_\rho(\gamma) + O\left(\frac{\log n}{n}\right)
\end{aligned}$$

Note that if $\gamma \in \mathcal{P}_N, \gamma \neq \rho$, then

$$I_\rho(\gamma) > 0$$

Therefore

$$P_n\{L_n \in B(\gamma,\varepsilon)\} \approx e^{-nI_\rho(\gamma)} \to 0 \qquad \text{as } n\to\infty$$

which is the law of large number:

$$\begin{aligned}
\lim_{n\to\infty} P_n\{L_n \in B(\gamma,\varepsilon)\} &= 0 \\
\lim_{n\to\infty} P_n\{L_n \in B(\rho,\varepsilon)\} &= 1
\end{aligned}$$

Furthermore, it gives the exponentially fast rate of decay. More generally, if A is a Borel subsets of $\mathcal{P}_N$ and if $\rho \notin$ the closure of A, then

$$\lim_{n\to\infty} P_n\{L_n \in A\} = 0$$

Notation

$$I_\rho(A) \equiv \inf_{\gamma\in A} I_\rho(\gamma)$$

then

$$\begin{aligned}
P_n\{L_n \in A\} &= \sum_{\gamma\in A} P_n\{L_n \sim \gamma\} \\
&\approx \sum_{\gamma\in A} e^{-nI_\rho(\gamma)}
\end{aligned}$$

Since

$$e^{-nI_\rho(A)} \leq \sum_{\gamma\in A} e^{-nI_\rho(\gamma)} \leq n^N e^{-nI_\rho(A)}$$

where, in the summation, L_n has the form of $\frac{k}{n}$, $k \in \mathbb{Z}^N$. Therefore, to exponential order, we have

$$P_n \{L_n \in A\} \approx e^{-nI_\rho(A)} \quad \text{as } n \to \infty$$

Now back to the conditional limit theorem: Let

$$A(z) \equiv \left\{ \gamma \in \mathcal{P}_N : \sum_k^N y_k \gamma_k \in [z-\delta, z] \right\}$$

i.e., the set of measure under which the mean is in $[z-\delta, z]$.

Note that

1. $\bar{\rho} \in A(z)$
2. Since $\frac{1}{n} S_n(\omega) = \sum_{k=1}^N y_k L_n(\omega, y_k)$, we have

$$\left\{ \omega \in \Omega_n : \frac{S_n(\omega)}{n} \in [z-\delta, z] \right\} = \{\omega \in \Omega_n : L_n(\omega) \in A(z)\}$$

$$\begin{aligned} P^* &\equiv P_n \left\{ L_n \in B(\bar{\rho}, \varepsilon) : \frac{S_n(\omega)}{n} \in [z-\delta, z] \right\} \\ &= P_n \{L_n \in B(\bar{\rho}, \varepsilon) : L_n(\omega) \in A(z)\} \\ &= \frac{P_n \{L_n \in B(\bar{\rho}, \varepsilon) \cap A(z)\}}{P_n \{L_n \in A(z)\}} \\ &\approx \frac{\exp[-nI_\rho(B(\bar{\rho}, \varepsilon) \cap A(z))]}{\exp[-nI_\rho(A(z))]} \\ &= e^{-n[I_\rho(B(\bar{\rho},\varepsilon)\cap A(z)) - I_\rho(A(z))]} \end{aligned}$$

Obviously,

$$I_\rho(B(\bar{\rho}, \varepsilon) \cap A(z)) \geq I_\rho(A(z))$$

Hence

$$P^* \sim O(1) \qquad \text{if } I_\rho(B(\bar{\rho}, \varepsilon) \cap A(z)) = I_\rho(A(z))$$

which is part 1 of the theorem. This is indeed the case as shown in the lemma:

Lemma: I_ρ attains its inf over $A(z)$ at the unique $\bar{\rho} = \{\rho_k\}$,

$$\bar{\rho} = \frac{\rho_k e^{-\beta y_k}}{\sum_{i=1}^N \rho_i e^{-\beta y_i}}$$

Proof: Since $\beta = \beta(z) > 0$,

$$Z[-\beta] = \log \sum_{k=1}^N \rho_k e^{-\beta y_k}$$

Note that

$$\frac{\bar{\rho}}{\rho_k} = \frac{e^{-\beta y_k}}{e^{Z[-\beta]}}$$

$\forall \gamma \in A(z)$

$$\begin{aligned} I_\rho(\gamma) &= \sum^N \gamma_k \log \frac{\gamma_k}{\rho_k} = \sum_{k=1}^N \gamma_k \log \frac{\gamma_k}{\bar{\rho}_k} + \sum_{k=1}^N \gamma_k \log \frac{\bar{\rho}_k}{\rho_k} \\ &= I_{\bar{\rho}}(\gamma) - \beta \sum_{k=1}^N \gamma_k y_k - Z[-\beta] \end{aligned}$$

for the second term of which we know $\gamma \in A(z)$, therefore,

$$\begin{aligned} I_\rho(\gamma) &\geq I_{\bar{\rho}}(\gamma) - \beta z - Z[-\beta] \\ &\geq -\beta z - Z[-\beta] \qquad (\because \quad I_{\bar{\rho}}(\gamma) \geq 0, \text{ "=" holds iff } \gamma = \bar{\rho}) \\ &= I_\rho(\bar{\rho}) \end{aligned}$$

which can be seen from

$$\begin{aligned} I_\rho(\bar{\rho}) &= \sum \bar{\rho}_k \log \frac{\bar{\rho}_k}{\rho_k} \\ &= \sum \bar{\rho}_k \log \left(\frac{\rho_k e^{-\beta y_k}}{\rho_k \sum_{i=1}^N \rho_i e^{-\beta y_i}} \right) \\ &= -\beta z - \log \sum_{k=1}^N \rho_k e^{-\beta y_k} \end{aligned}$$

Hence, we have

$$I_\rho(\gamma) \geq I_\rho(\bar{\rho})$$

and the equality holds if and only if $\gamma = \bar{\rho}$.

To prove part 3 of the theorem:

$\forall$function $\varphi : \Lambda \to \mathbb{R}$, we define a continuous function on $\mathcal{P}_N$ via

$$f(\gamma) \equiv \sum_{k=1}^N \varphi(y_k) \gamma_k$$

Therefore,

$$\begin{aligned} f(L_n) &= \sum_{k=1}^N \varphi(y_k) L_n(y_k) \\ &= \frac{1}{n} \sum_{j=1}^n \varphi(X_j) \end{aligned}$$

$$\begin{aligned}
\lim_{n\to\infty} \mathbf{E}_{P_n}\left\{\varphi(X_1) : \frac{S_n}{n} \in [z-\delta, z]\right\} &= \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} \mathbf{E}_{P_n}\left\{\varphi(X_j) : \frac{S_n}{n} \in [z-\delta, z]\right\} \quad \text{(by symmetry)} \\
&= \lim_{n\to\infty} \mathbf{E}_{P_n}\left\{f(L_n) : \frac{S_n}{n} \in [z-\delta, z]\right\} \\
&= f(\bar{\rho}) \quad \text{(using Part 2)} \\
&= \sum_{k=1}^{N} \varphi(y_k)\,\bar{\rho}_k
\end{aligned}$$

Choose $\varphi = I_{y_k}$, then

$$\lim_{n\to\infty} P_n\left\{X_1 = y_k : \frac{S_n}{n} \in [z-\delta, z]\right\} = \bar{\rho}_k$$

QED.

1.2.2 Maximum Entropy Principle

As discussed above, conditioned on $\frac{S_n}{n} \in [z-\delta, z]$, the asymptotically most probable configuration of L_n is $\bar{\rho}$. If we define

$$S(\gamma, \rho) \equiv -I_\rho(\gamma),$$

then, $\bar{\rho}$ is uniquely determined by maximizing $S(\gamma, \rho)$, i.e.,

$$\min_{\gamma\in A(z)} I_\rho(\gamma) = \max S(\gamma, \rho).$$

Clearly, were not for the condition, $\gamma \in A(z)$, we would have

$$\beta = 0, \quad \bar{\rho}_k = \rho_k$$

then,

$$P_n[L_n \in B(\gamma, \varepsilon)] \approx e^{-nI_\rho(\gamma)} \qquad \text{as} \;\; n \to \infty$$

The **Maximum Entropy Principle** is as follows:

$\gamma_0 \in \mathcal{P}_N$ is an equilibrium value of L_n with respect to P_n if and only if $\min_{\mathcal{P}_N} I_\rho(\gamma)$ yields $\gamma_0 = \rho$

1.2.3 An Application — Discrete Ideal Gas

Consider n identical, noninteracting particles, each having N equally likely energy levels $y_1, \cdots, y_N$. Clearly, in this case, we have

$$\begin{aligned} \Lambda &= \{y_1, \cdots, y_N\} \\ \rho_k &= \frac{1}{N} \end{aligned}$$

A possible configuration, ω, of the gas is an element of $\Omega_n = \Lambda^n$, the total energy of the configuration ω is

$$E_n(\omega) = \sum_{j=1}^{n} \omega_j = S_n(\omega)$$

Suppose our measurement of the average energy per particle gives approximately $\frac{E_n}{n} \approx z$. More precisely,

$$\frac{E_n(\omega)}{n} \in [z-\delta, z]$$

where $y_1 \leq z-\delta < z < \bar{y}$, where

$$\bar{y} = \sum_{k=1}^{N} y_k \rho_k$$

The asymptotically most likely probability of a particle occupying the energy level k is

$$\bar{\rho}_k = \lim_{n\to\infty} P_n \left\{ X_1 = y_k : \frac{E_n(\omega)}{n} \in [z-\delta, z] \right\}$$

According to our conditional limit theorem above, this most probably probability is given by the maximum entropy principle, i.e.,

$$\bar{\rho}_k = \frac{e^{-\beta y_k}}{\sum_i e^{-\beta y_i}}$$

where the parameter β is determined by

$$\sum_{k=1}^{N} y_k \bar{\rho}_k = z$$

Homework:

1. For the discrete gas, consider the situation where the measurement yields $E_n/n \in [z-\delta, z] \subset (\bar{y}, y_N]$. Is the parameter β positive for this case?

2. Find the most probable measure $\rho = \rho^*$ on $\Lambda = \{y_1, \cdots, y_N\}$ that is consistent with the following conditions:

(a) a finite number of statistical measurements F_j of given function $f_{j,}$ $j = 1, \cdots, r$. $r \leq N-1$, i.e.,

$$F_j = \langle f_j \rangle_p = \sum_{k=1}^{N} f_j(y_k)\, p_k,\ \ j = 1, \cdots, r$$

and

(b) the external bias given by ρ^0.

Solution for problem 2: We have to solve the optimization problem:

$$\max_{\substack{F_j = \langle f_j \rangle, j=1,\cdots,r \\ \sum_{k=1}^{N} p_k = 1}} S\left(\rho, \rho^0\right)$$

In the absence of the constraints, $F_j = \langle f_j \rangle_p$, the optimization would yields

$$\rho^* = \rho^0$$

as our external bias. With the constraints, we invoke the Lagrange multipliers, which yields

$$\left\{ \partial_\rho S\left(\rho, \rho^0\right) - \sum_{j=1}^{r} \beta_j \partial_\rho \langle f_j \rangle_\rho - \beta_0 \partial_\rho \sum_{k=1}^{N} \rho_k \right\}\Bigg|_{\rho=\rho^*} = 0$$

Thus, we have

$$\ln \frac{\rho_k}{\rho_k^0} = -\sum_{j=1}^{r} \beta_j f_j(y_k) - (\beta_0 + 1),\ \ 1 \leq k \leq N$$

therefore,

$$\rho_k^* = \rho_k^0 e^{-\sum_{j=1}^{r} \beta_j f_j(y_k)} e^{-(\beta_0+1)} \tag{5}$$

Since

$$\sum_{k=1}^{N} \rho_k^* = 1,$$

Eq. (5) leads to

$$e^{-(\beta_0+1)} = \sum_{k=1}^{N} \rho_k^0 e^{-\sum_{j=1}^{r} \beta_j f_j(y_k)}$$

thus

$$\rho_k^* = \frac{\rho_k^0 e^{-\sum_{j=1}^{r} \beta_j f_j(y_k)}}{\sum_{k'=1}^{N} \rho_{k'}^0 e^{-\sum_{j=1}^{r} \beta_j f_j(y_{k'})}},\quad k = 1, \cdots, N$$

The corresponding partition function is

$$Z\left(\left\{\beta_j\right\}, \rho_0\right) \equiv \log\left[\sum_{k=1}^{N} \rho_k^0 e^{-\sum_{j=1}^{r} \beta_j f_j(y_k)}\right]$$

where the Lagrange multipliers satisfy

$$\langle f_j \rangle_{\rho^*} = -\frac{\partial}{\partial \beta_j} Z\left(\{\beta_j\}, \rho_0\right).$$

2 Canonical Ensemble:

2.1 Liouville Property and Liouville's Theorem

If a system is described by

$$\begin{cases} \frac{d}{dt}\mathbf{X} = \mathbf{F}(\mathbf{X}), & \mathbf{X} \in \mathbb{R}^N \\ \mathbf{X}|_{t=0} = \mathbf{X}_0 \end{cases}$$

Liouville property:

$$\nabla_{\mathbf{x}} \cdot \mathbf{F} = \sum_{j=1}^{N} \frac{\partial F_j}{\partial X_j} = 0$$

the corresponding flow map Φ^t :

$$\mathbb{R}^N \longmapsto \mathbb{R}^N$$

$$\begin{cases} \frac{d}{dt}\Phi^t(\mathbf{X}) = \mathbf{F}\left(\Phi^t(\mathbf{X})\right), & \mathbf{X} \in \mathbb{R}^N \\ \Phi^t(\mathbf{X})|_{t=0} = \mathbf{X}_0 \end{cases}$$

Then,

1. $\det\left(\nabla_{\mathbf{x}}\Phi^t(\mathbf{X})\right) = 1$, i.e., volume preserving.
2. If $p(\mathbf{X},t) \equiv p_0\left(\left(\Phi^t\right)^{-1}(\mathbf{X})\right)$, $p_0(\mathbf{X})$ is the initial probability density function at $t = 0$, then
$$\frac{\partial}{\partial t}p + \mathbf{F} \cdot \boldsymbol{\nabla}_{\mathbf{X}} p = 0$$
3. If $G(p)$ is any function of the pdf p, then
$$\frac{\partial}{\partial t}G(p) + \mathbf{F} \cdot \boldsymbol{\nabla}_{\mathbf{X}} G(p) = 0$$
i.e., $G(p)$ satisfies the Liouville's theorem. Therefore,
$$\frac{d}{dt}\int_{\mathbb{R}^N} G(p(\mathbf{X}(t)))\, d\mathbf{X} = 0$$

this can be seen as follows:

$$
\begin{aligned}
\frac{\partial}{\partial t}G(p)+\mathbf{F}\cdot\boldsymbol{\nabla}_{\mathbf{X}}G(p) &= G'(p)\left(\frac{\partial}{\partial t}p+\mathbf{F}\cdot\boldsymbol{\nabla}_{\mathbf{X}}p\right)=0\\
\frac{d}{dt}\int_{\mathbb{R}^N}G(p(\mathbf{X}(t)))\,d\mathbf{X} &= \int_{\mathbb{R}^N}\frac{\partial}{\partial t}G(p)\,d\mathbf{X}\\
&= -\int_{\mathbb{R}^N}\mathbf{F}\cdot\boldsymbol{\nabla}_{\mathbf{X}}G(p)\,d\mathbf{X}\\
&= -\int\boldsymbol{\nabla}_{\mathbf{X}}\cdot[G(p)\,\mathbf{F}]\,d\mathbf{X} \qquad (\because\ \ \nabla_{\mathbf{x}}\cdot\mathbf{F}=0,\ \text{i.e., Liouville Property})\\
&= 0 \qquad \text{(Vanishing boundary conditions)}
\end{aligned}
$$

2.2 Conservation Laws

Suppose there exist L conserved quantities $E_l(\mathbf{X}(t))$, i.e.,

$$E_l(\mathbf{X}(t))=E_l(\mathbf{X}_0)\,, \quad 1\le l\le L$$

The average with respect to the pdf p is defined by

$$\langle E_l\rangle_p\equiv\int_{\mathbb{R}^N}E_l(\mathbf{X})\,p(\mathbf{X})\,d\mathbf{X}$$

It can be easily seen that the averages are conserved in time, that is,

$$\langle E_l\rangle_{p(\mathbf{X},t)}=\langle E\rangle_{p_0(\mathbf{X})}\,, \quad \forall t$$

since

$$
\begin{aligned}
\langle E_l\rangle_{p(\mathbf{X},t)} &= \int_{\mathbb{R}^N}E_l(\mathbf{X})\,p(\mathbf{X},t)\,d\mathbf{X}\\
&= \int_{\mathbb{R}^N}E_l(\mathbf{X})\,p_0\left(\left(\Phi^t\right)^{-1}(\mathbf{X})\right)d\mathbf{X}\\
&\underset{\mathbf{Y}=(\Phi^t)^{-1}\mathbf{X}}{=} \int_{\mathbb{R}^N}E_l\left(\Phi^t(\mathbf{Y})\right)p_0(\mathbf{Y})\,d\mathbf{Y} \quad \text{(Volume preserving under } \Phi)\\
&= \int_{\mathbb{R}^N}E_l(\mathbf{Y})\,p_0(\mathbf{Y})\,d\mathbf{Y} \qquad (E_l\ \text{ is conserved})\\
&= \langle E_l\rangle_{p_0}
\end{aligned}
$$

2.3 Entropy and Maximum Entropy Principle

Entropy is defined as

$$S(p) \equiv -\int_{\mathbb{R}^N} p(\mathbf{X}) \ln p(\mathbf{X})\, d\mathbf{X}$$

Clearly, $S(p)$ is conserved in time since, for any function G of p, $G(p)$ satisfies

$$\frac{d}{dt}\int G(p)\, d\mathbf{X} = 0$$

The question is which p one should use for describing statistical ensembles. From the large deviation principle, we know that it should be a pdf such that it satisfies the maximum entropy principle, i.e., p^* for the most probable state, such that

$$S(p^*) = \max_{p \in C} S(p)$$

where C is the set of constraints:

$$C = \left\{ p(\mathbf{X}) \geq 0, \quad \int_{\mathbb{R}^N} p(\mathbf{X})\, d\mathbf{X} = 1, \quad \langle E_l \rangle_p = E_l, \quad 1 \leq l \leq L \right\}.$$

2.4 The Most Probable State and Gibbs Measure

Using Lagrange multipliers $\beta_0, \beta_1, \cdots \beta_L$, we maximize

$$S(p) - \sum_{l=1}^{L} \beta_l \left(\langle E_l \rangle_p - E_l \right) - \beta_0 \left(\int p\, d\mathbf{X} - 1 \right).$$

$$\frac{\delta S(p)}{\delta p(\mathbf{X})} = -\int \frac{\delta p(\mathbf{X}')}{\delta p(\mathbf{X})} \ln p(\mathbf{X}')\, d\mathbf{X} - \int p(\mathbf{X}') \frac{\partial \ln p(\mathbf{X}')}{\partial p(\mathbf{X})} \frac{\delta p(\mathbf{X}')}{\delta p(\mathbf{X})} d\mathbf{X}'$$

Since

$$\frac{\delta p(\mathbf{X}')}{\delta p(\mathbf{X})} = \delta(\mathbf{X} - \mathbf{X}') \text{ — Dirac } \delta\text{-function}$$

$$\therefore \quad \frac{\delta S(p)}{\delta p(\mathbf{X})} = -\ln p(\mathbf{X}) - 1$$

Furthermore

$$\begin{aligned} \frac{\delta \langle E_l \rangle_p}{\delta p(\mathbf{X})} &= \int E_l(\mathbf{X}') \frac{\delta p(\mathbf{X}')}{\delta p(\mathbf{X})} d\mathbf{X}' \\ &= E_l(\mathbf{X}) \end{aligned}$$

therefore,

$$
\begin{aligned}
-\left(1+\ln p^*\right) &= \beta_0 + \sum_{l=1}^{L} \beta_l E_l\left(\mathbf{X}\right) \\
\therefore \quad p^*\left(\mathbf{X}\right) &= \mathcal{N} \exp\left[-\sum_{l=1}^{L} \beta_l E_l\left(\mathbf{X}\right)\right]
\end{aligned}
$$

where $\mathcal{N}$ is a normalization factor. This is precisely the Gibbs measure for canonical ensembles.

Hence, a Gibbs measure describes the most probable state in the sense of maximum entropy with constraints.

Note that

1. Gibbs measure solves the steady Liouville's equation. This is merely a special case of the following general theorem.

   Theorem: If $E_l,\ 1 \leq l \leq L$, are conserved under the evolution:

   $$
   \begin{cases}
   \frac{d}{dt}\mathbf{X} = \mathbf{F}\left(\mathbf{X}\right), & \mathbf{X} \in \mathbb{R}^N \\
   \mathbf{X}|_{t=0} = \mathbf{X}_0, &
   \end{cases}
   $$

   For any smooth function $g\left(E_1, \cdots, E_L\right)$, $g\left(E_1, \cdots, E_L\right)$ satisfies

   $$\mathbf{F} \cdot \boldsymbol{\nabla}_{\mathbf{X}} g = 0$$

   This is obvious since

   $$
   \begin{aligned}
   & E_l\left(\mathbf{X}\left(t\right)\right) \text{ is conserved in time,} \\
   \therefore \quad & 0 = \frac{d}{dt}E_l\left(\mathbf{X}\left(t\right)\right) = \frac{\partial E_l}{\partial t} + \mathbf{F} \cdot \boldsymbol{\nabla}_{\mathbf{X}} E_l = \mathbf{F} \cdot \boldsymbol{\nabla}_{\mathbf{X}} E_l, \ \forall l \\
   \therefore \quad & \mathbf{F} \cdot \boldsymbol{\nabla}_{\mathbf{X}} g\left(E_1, \cdots, E_L\right) = \sum_{l=1}^{L} \mathbf{F} \cdot \boldsymbol{\nabla}_{\mathbf{X}} E_l \frac{\partial g}{\partial E_l} = 0
   \end{aligned}
   $$

2. Gibbs measure is an invariant measure.

   What is an invariant measure? A measure μ on $\mathbb{R}^N$ is invariant under the flow map Φ^t if

   $$\mu\left(\left(\Phi^t\right)^{-1}\left(\Omega\right)\right) = \mu\left(\Omega\right) \quad \forall t$$

   for any measurable set $\Omega \subset \mathbb{R}^N$. The invariance of the Gibbs measure now can be seen as follows:

$$
\begin{aligned}
&\frac{d}{dt}\int_{(\Phi^t)^{-1}\Omega} p^*(\mathbf{X})\,d\mathbf{X} \\
= \;& \frac{d}{dt}\int_{\Omega} p^*\left(\left(\Phi^t\right)^{-1}(\mathbf{Y})\right)d\mathbf{Y} \qquad \left(\mathbf{Y} = \Phi^t(\mathbf{X}) \text{ and volume preserving}\right) \\
= \;& \int_{\Omega} \mathbf{F}\cdot\boldsymbol{\nabla}_{\mathbf{X}} p^*(\mathbf{X})\,|_{\mathbf{X}=(\Phi^t)^{-1}(\mathbf{Y})} d\mathbf{Y} \\
= \;& 0 \\
\therefore \;& \int_{(\Phi^t)^{-1}\Omega} p^*(\mathbf{X})\,d\mathbf{X} \text{ is independent of time } t \\
\text{i.e., } \int_{(\Phi^t)^{-1}\Omega} p^*(\mathbf{X})\,d\mathbf{X} = \;& \int_{\Omega} p^*(\mathbf{X})\,d\mathbf{X} \\
& \text{— Gibbs measure is an invariant measure of } \Phi^t.
\end{aligned}
$$

3. Relation between Energy and Entropy under the Gibbs measure:

$$
\left.\frac{\delta S(p(\mathbf{X}))}{\delta\langle E(\mathbf{X})\rangle}\right|_{p^*} = \left.\frac{\frac{\delta S(p(\mathbf{X}))}{\delta p(\mathbf{X})}}{\frac{\delta\langle E(\mathbf{X})\rangle}{\delta p(\mathbf{X})}}\right|_{p^*} = \frac{-\ln p^*(\mathbf{X}) - 1}{E(\mathbf{X})} \tag{6}
$$

Using

$$
p^*(\mathbf{X}) = \frac{e^{-\beta E(\mathbf{X})}}{\int e^{-\beta E(\mathbf{X})} d\mathbf{X}}
$$

then, Eq. (6) becomes

$$
\begin{aligned}
\left.\frac{\delta S(p(\mathbf{X}))}{\delta\langle E(\mathbf{X})\rangle}\right|_{p^*} &= \frac{\beta E(\mathbf{X}) + \beta_0}{E(\mathbf{X})} \\
&\longrightarrow \beta
\end{aligned}
$$

in the thermodynamic limit ($N \to \infty \Longrightarrow E(\mathbf{X}) \to \infty$).

Note Title 10/10/2006

idea gas : $p = \frac{N}{V} k_B T$

In general, Virial expansion

$$p = \left(\frac{N k_B T}{V}\right)\left[1 + \frac{N}{V} B_2(T) + \left(\frac{N}{V}\right)^2 B_3(T) + \cdots\right]$$

determined by intermolecular force

First law: Conservation of Energy

$$dU = đQ - đW$$

heat work

e.g. $đW = p\,dV - \vec{E}\,d\vec{P} - \vec{H}\cdot d\vec{M} - \mu\,dN$

$-p, \vec{E}, \vec{H}, \mu$ — generalized force Y

$dV\ d\vec{P}\ d\vec{M}\ dN$ — generalized displacement X

$\therefore \quad dU = dQ + Y\,dX + \mu\,dN$

(2nd law): $dQ \le T\,dS$ "=" reversible process

For reversible process $dU = T\,dS + Y\,dX + \mu\,dN$

in general $dU \le T\,dS + Y\,dX + \mu\,dN$

Fundamental Equation of Thermodynamics

Thermodynamically, entropy — extensive.

i.e additive wrt independent subsystems

$\therefore$ $$S(\lambda U, \lambda X, \lambda N) = \lambda S(U, X, N)$$

i.e. 1st-order homogeneous fn of the extensive state variable

NB. $$\frac{d}{d\lambda}(\lambda S) = \left(\frac{\partial S}{\partial(\lambda U)}\right)\frac{d}{d\lambda}(\lambda U) + \left(\frac{\partial S}{\partial(\lambda X)}\right)\frac{d}{d\lambda}(\lambda X) + \left(\frac{\partial S}{\partial(\lambda N)}\right)\left(\frac{d}{d\lambda}\lambda N\right) \qquad (*)$$

$\because$ $$dU = TdS + YdX + \mu dN$$

or $$dS = \frac{1}{T}dU - \frac{Y}{T}dX - \frac{\mu}{T}dN$$

$\therefore$ $$\left(\frac{\partial S}{\partial U}\right) = \frac{1}{T}, \quad \left(\frac{\partial S}{\partial X}\right) = -\frac{Y}{T}, \quad \left(\frac{\partial S}{\partial N}\right) = -\frac{\mu}{T}$$

eq of state: thermal, mechanical, chemical

Therefore: Eq $(*) \Rightarrow$

$$TS = U - YX - \mu N \qquad (**)$$

— the fundamental eq of thermodynamics

The Gibbs-Duhem eqn:

$$d(**) \Rightarrow \quad \underbrace{TdS}_{= TdS + YdX + \mu dN} + SdT = dU - YdX - XdY - \mu dN - Nd\mu$$

$$\therefore \quad SdT + XdY + \mu dN = 0 \quad \text{(G-D eqn)}$$

Helmholtz Free Energy.

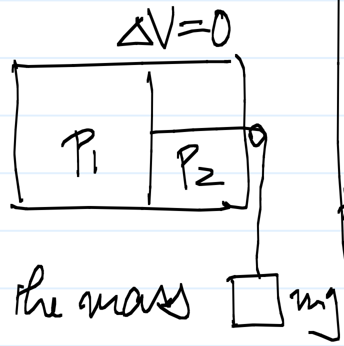
System: closed and thermally coupled to a heat bath.
mechanical isolated (i.e. V is constant)

$$A = U - ST = YX + \mu N$$

$$\therefore dA = dU - SdT - TdS$$

$$\downarrow \; dU = dQ - dW \le TdS - dW$$

Why free energy?



NB: if isolated, heat-isolating box, then $(\Delta U)_{S,V,N} \le -\Delta W_{free}$

if $P_1 a + mg > P_2 a$: (a = area)
work done on gas by the mass

if $P_1 a + mg < P_2 a$, $\Rightarrow$ work done on the mass by the gas

$$dW = \int p\,dV + \Delta W_{free}$$

$$\therefore \Delta A \le -\int S\,dT - \int p\,dV - \Delta W_{free} + \int \mu\,dN$$

For fixed V, N, T. $\therefore (\Delta A)_{T,N,V} \le -\Delta W_{free}$

For a reversible process at constant T, X, and N, V work that can be stored and recovered completely is the Helmholtz free energy.

If no work is done for a process of fixed T, V, N,

then $(\Delta A)_{V,T,N} \leq 0$

NB. an equilibrium state cannot be changed spontaneously,
Helmholtz free energy can only decrease or not changing,
$\therefore$ the equilibrium state is a min free energy state.

2.5 Thermodynamic Relations

Physically, entropy is defined with a unit, i.e.,

$$S = -k_B \int_{\mathbb{R}^N} p(\mathbf{X}) \ln(\mathcal{N} p(\mathbf{X}))\, d\mathbf{X}$$

where, e.g., $d\mathbf{X} = d^{3N}p d^{3N}q$ and

$$\mathcal{N} = \begin{cases} N!\,(2\pi\hbar)^{3N} & \text{for indistinguishable particles} \\ (2\pi\hbar)^{3N} & \text{for distinguishable particles} \end{cases}$$

2.5.1 Partition function and Free Energy

The partition function is defined as

$$Q_N(V,T) \equiv \int \frac{d^{3N}p d^{3N}q}{\mathcal{N}} \quad e^{-\beta H(p,q)}$$

where $H(p,q)$ is the Hamiltonian of a system in interest and

$$\beta = \frac{1}{k_B T}$$

Helmholtz free energy is defined as

$$A(V,T) \equiv -k_B T \log Q_N(V,T)$$

Note that

$$A = U - TS$$

where $U = \langle H \rangle$, which can be seen as follows:

$$\begin{aligned} \because \quad & Q_N = e^{-\beta A} \\ \therefore \quad & \int \frac{d^{3N}p d^{3N}q}{\mathcal{N}} \quad e^{-\beta[A(V,T)-H(p,q)]} = 1 \end{aligned}$$

$\frac{\partial}{\partial \beta} :\Longrightarrow$

$$\int \frac{d^{3N}p d^{3N}q}{\mathcal{N}} \quad e^{-\beta[A(V,T)-H(p,q)]} \left[A(V,T) - H(p,q) + \beta \left(\frac{\partial A}{\partial \beta} \right)_V \right] = 0$$

i.e.,

$$A(V,T) - U(V,T) + k_B T \left[\beta^2 \left(\frac{\partial A}{\partial \beta} \right)_V \right] = 0 \tag{7}$$

Since

$$A = -\frac{1}{\beta} \ln Q_N = -\frac{1}{\beta} \ln \int d\mathbf{X} e^{-\beta H}$$

(where, for simplicity, we have set $\mathcal{N} = 1$)

$$\beta^2 \left(\frac{\partial A}{\partial \beta}\right)_V = \frac{\beta \int d\mathbf{X} H e^{-\beta H}}{\int d\mathbf{X} e^{-\beta H}} + \ln Q_N,$$

Furthermore,

$$\begin{aligned} S\left(p^*\left(\mathbf{X}\right)\right) &= -k_B \int \frac{e^{-\beta H}}{Q_N} \ln\left(\frac{e^{-\beta H}}{Q_N}\right) d\mathbf{X} \\ &= -\frac{k_B}{Q_N} \int e^{-\beta H} \left[\left(-\beta H\right) - \ln Q_N\right] d\mathbf{X} \\ &= k_B \beta \frac{\int d\mathbf{X} H e^{-\beta H}}{Q_N} + k_B \ln Q_N \end{aligned}$$

therefore,

$$S = k_B \beta^2 \left(\frac{\partial A}{\partial \beta}\right)_V$$

Thus, from Eq. (7) we have

$$A = U - TS$$

In statistical physics, the pressure is defined by

$$\begin{aligned} P &\equiv -\left(\frac{\partial A}{\partial V}\right)_T \\ &= \frac{\partial}{\partial V}\left(k_B T \ln Q_N\left(V, T\right)\right) \end{aligned}$$

which gives the equation of state.

The first law of thermodynamics, which is

$$dU = TdS - PdV$$

which can be shown as follows:

$$\begin{aligned} &\because \quad A = U - TS \\ &\therefore \quad dA = dU - SdT - TdS \end{aligned} \tag{8}$$

Moreover

$$\begin{aligned} dA &= \frac{\partial A}{\partial V} dV + \frac{\partial A}{\partial T} dT \\ &= -PdV - SdT \end{aligned} \tag{9}$$

where use is made of

$$\begin{aligned} S &= k_B \beta^2 \left(\frac{\partial A}{\partial \beta}\right)_V \\ &= -\left(\frac{\partial A}{\partial T}\right)_V \end{aligned}$$

Comparing Eq. (8) and (9) yields

$$dU = TdS - PdV.$$

Homework: Use canonical ensemble to derive the equation of state for the ideal gas and evaluate the entropy.