Statistical Mechanics

1 Priori

1.1 Probability

Probability is a fundamental information lost measure. Suppose we know the dynamics of a process, and its initial condition, we can deduce the deterministic results.

So, probability must be a measure upon some potential state set to reflect the state in **frequency** or **understanding** to such system. We say $P : \Omega \to \mathcal{B}$ with element measure as $P(d\mu)$ for $d\mu \in \Omega$, or discrete $P(\omega)$ for $\omega \in \Omega$. Thus, we can take the summon on such element due to disjointness that $P(\cup_i \omega_i) = P(A)$ where $\omega_i \in A$.

$$\frac{P(A \cap B)}{P(B)} = P(A|B) \tag{1.1}$$

Indicates that the conditional probability $A|B$ of a given presumed condition B. Suppose one has $P(A \cap B = \emptyset) = 0$, one has $P(A|B) = 0$. If $A \subseteq B \to A \cap B = A$, we has a simpler expression $P(A)/P(B)$, which is akin to a subset from whole. Another special case is $P(A|B) = P(A)$ where the condition is indifferent to A, where one has $P(A \cap B) = P(A)P(B)$.

$$P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A) \tag{1.2}$$

Then given that $\Omega = \cup_i A_i$ for disjoint partition, we has:

$$P(A \cap \Omega = A) = P(A|\Omega) \cdot P(\Omega) = \sum_i P(A|B_i)P(B_i) \tag{1.3}$$

Given a projection or random variable $X : \Omega \to \mathcal{B}$ where $\mathcal{B}$ is measurable set, which, you may want it's also a metric and separable space, however, the crucial thing is one commonly investigate the state set by such function, suppose we have $t \to f(t)$ in classical mechanics, we inspect that $f(t + T) = f(t)$, then we retroactively conclude that in such system, time is in period. The same thing happens upon the state space. We redistribute the state space through a coarse measure of it to acquire a dynamics quality. Such thing can be restated in statistical mechanics that we have many micro states of particles and a projection or macro state to **reflect** the system.

$$\int_\Omega P(d\mu) = 1_{\mathcal{B}} \tag{1.4}$$

Sometimes given a already measurable state space, we will notate also:

$$\int_\Omega P(d\mu) \equiv \int_\Omega p(\mu)d\mu \equiv \int_\Omega dP$$
$$\frac{dP}{d\mu} = p$$

(1.5)

Thus we can extend probability from state set to random variable as indicators:

$$P(X \in \mathcal{B}) = \int_{\{\omega \in \Omega | X(\omega) \in \mathcal{B}\}} P(d\mu)$$
$$\mathbb{E}(X) = \langle X \rangle = \int_\Omega X P(d\mu)$$

(1.6)

Thus if one has no additional investigation upon such state space, each state is indifferent and thus, **equal** in respective view of evaluation. Such **priori** probability is called uniform distribution, which is the basic results comes from indifference of elements of a set. How? A uneven distribution would reflect the underneath state dynamics by given intersections or **section** in local view. Then for even case, each $P(\cap_i^\infty A_i) \to P(\omega)$ must be equal for any set convergence, resulting the same local view.

$$P(\cup_I \omega_i) = \sum_i P(\omega_i)/\int_\Omega P(d\omega \in \Omega) = 1_\mathcal{B}$$
$$P(\omega_i) = \frac{1}{|\Omega|} \text{ or } P(\omega) = \delta(\mu = \omega)$$

(1.7)

I avoid $\delta(\mu - \omega)$ to adapt set without any natural number system. Thus, it′s a uniform distribution throughout whole isolated basic states.

A strong suggestion of such **priori** equal probability is information, which is the most unknown condition for observer. If you read few axioms about information and entropy origin, you will find that it′s a delicate suitable restriction to deduce the results form. For example, strong additivity:

$$H(X, Y)_p = H(X)_p + H(Y|X)_p$$
$$H(X, Y)_p = H(X)_p + H(Y)_p \quad \text{if } X, Y \text{ independent} \to H(Y|X) = H(Y)$$

(1.8)

For a given probability distribution. Given above maximal condition upon uniform distribution, we has below:

$$H(X)_p = -\int_X d\mu p(\mu) \log p(\mu) = -\int_X \log\left(\frac{dP}{d\mu}\right) dP$$
$$H(X)_{\frac{1}{N}} = \sum_i \frac{1}{N} \log N = \log N$$

(1.9)

Often, we call $-\log p(\mu)$ information $I(X)_p$. It is suitable for any restriction based on human imaginary, for example, you should acquire more information if the probability is low because you need to update your knowledge based on Bayesian!

$$H(X, Y)_p = -\int_{X,Y} d\mu p(x, y) \log p(x, y)$$

$$= -\int_{X,Y} d\mu p(x, y) \log p(x) p(y|x) \tag{1.10}$$

$$= -\int_X d\mu p(x) \log p(x) - \int_{X,Y} p(x, y) \log p(y|x)$$

$$H(Y|X) = \int_X d\mu p(x) H(Y|X = x)$$

$$\tag{1.11}$$

$$= -\int_{X,Y} d\mu p(x) p(y|x) \log p(y|x)$$

Where $H(X) = \mathbb{E}(I(X))$, thus fix a element x, one has $H(Y|X = x) = \mathbb{E}(I(Y|X = x))$. Then we integrate or sum on the state subset X.

$$S(p_i, \lambda_a) = -\sum_i p_i \log p_i + \sum_a \lambda_a \left(\sum_i X_i^a p_i - X^a \right) \tag{1.12}$$

With λ_a the Lagrange multiplier, $\sum_i X_i^a = X^a$ the summation constraint. We has such extreme point:

$$\frac{\partial S}{\partial p_i} = -\log p_i - 1 + \sum_a \lambda_a X_i^a = 0$$

$$\tag{1.13}$$

$$\frac{\partial S}{\partial \lambda_a} = \sum_i X_i^a p_i - X^a = 0$$

Thus we have the distribution:

$$p_i = \exp\left(\sum_a \lambda_a X_i^a - 1 \right) = \frac{1}{Z} \exp\left(\sum_a \lambda_a X_i^a \right) \tag{1.14}$$

We can extend such in continuous case by the generalized Lebesgue integral, therefore a example can be shown that:

$$\int d\mu(x) P(x) = 1 \quad \int d\mu(x) x P(x) = \langle x \rangle \quad \int d\mu(x) P(x) x^2 = \langle x^2 \rangle \tag{1.15}$$

$$P(x) = C e^{\lambda_1 x + \lambda_2 x^2}$$

$$= C \exp\left(\frac{\lambda_1^2}{4\lambda_2} \right) \exp\left(-\lambda_2 \left(x + \frac{\lambda_1}{2\lambda_2} \right)^2 \right) \tag{1.16}$$

$$= A \exp(...)$$

Now we try to deduce coefficients λ_1 and λ_2 by the constraint:

$$x \to x + \frac{\lambda_1}{2\lambda_2} = y$$

$$\int_{-\infty}^{\infty} P(x)dx = A\sqrt{\frac{\pi}{\lambda_2}} = 1$$

$$A\int_{-\infty}^{\infty} \left(y - \frac{\lambda_1}{2\lambda_2}\right)e^{-\lambda_2 y^2}dy = -A\frac{\lambda_1}{2\lambda_2}\sqrt{\frac{\pi}{\lambda_2}} = \langle x \rangle \tag{1.17}$$

$$A\int_{-\infty}^{\infty} \left(y - \frac{\lambda_1}{2\lambda_2}\right)^2 e^{-\lambda_2 y^2}dy = \frac{1}{2\lambda_2} + \frac{\lambda_1^2}{4\lambda_2^2} = \langle x^2 \rangle$$

$$\frac{\lambda_1}{\lambda_2} = -2\langle x \rangle$$

$$\frac{1}{2\lambda_2} = \langle x^2 \rangle - \langle x \rangle^2 \tag{1.18}$$

$$\langle x \rangle = \alpha \quad \langle x^2 \rangle - \langle x \rangle^2 = \beta^2$$

$$P(x) = A\exp\left(\frac{(x-\alpha)^2}{2\beta^2}\right) \tag{1.19}$$

Is the final answer, which is a Gaussian distribution with mean α and variance β^2.

You can deduce the maximal entropy by Lagrange multiplier with restriction $\sum_i p_i = 1$, which is coincide to the uniform distribution. However, you can't deduce reversely by such condition, because there are a huge set of functional suitable for uniform convergence, for example $H^? = \sum_i p_i^2$.

$$S_n = \frac{1}{n}\sum_i^N X_i \geq x \tag{1.20}$$

For non-negative random variable space X:

$$\mathbb{E}(X) = \int_X X dP(X) \geq \int_{X \geq a} X dP(X) \geq a\int_{X \geq a} dP(X) = aP(X \geq a) \tag{1.21}$$

$$\mathbb{E}(X) \geq aP(X \geq a) \tag{1.22}$$

$$\mu_n = \mathcal{B}\left(\frac{1}{n}\sum_i^N X_i\right) = P_{\frac{1}{n}\sum_i^N X_i} \tag{1.23}$$

4

$$\mu_n([x, \infty)) = P\left(\frac{1}{n}\sum_i^N X_i \geq x\right)$$

$$= P\left(\exp\left(\lambda \sum_i^n X_i\right) \geq \exp(\lambda n x)\right)$$

$$\leq \exp(-\lambda n x)\mathbb{E}\left(\exp\left(\lambda \sum_i^N X_i\right)\right) \tag{1.24}$$

$$= \exp(-n(\lambda x - \Lambda(\lambda))) \quad \Lambda = \ln\mathbb{E}\left(\exp\left(\lambda \sum_i^N X_i\right)\right)$$

$$\Lambda^*(\lambda, x) = \lambda x - \Lambda(\lambda) \tag{1.25}$$

We could explicitly evaluate on.

$$P(S_n = s) = \int_{\{S_n(\mu)=s\}} p(\mu)d\mu = \int_\Omega \delta(S_n(\mu) - s)p(\mu)d\mu = \langle\delta(S_n(\mu) - s)\rangle \tag{1.26}$$

$$\delta(s) = \frac{1}{2\pi}\int e^{i\zeta s}d\zeta = \frac{1}{2\pi i}\int_{-i\infty+a}^{i\infty+a} e^{\zeta s}d\zeta \tag{1.27}$$

$$P(S_n = s) = \frac{1}{2\pi i}\int_{-i\infty+a}^{i\infty+a} d\zeta \int_\Omega d\mu p(\mu)e^{\zeta S_n(\mu)-s} = \frac{1}{2\pi i}\int_{-i\infty+a}^{i\infty+a} d\zeta e^{-\zeta s}\int_\Omega d\mu\, p(\mu)e^{\zeta S_n(\mu)} \tag{1.28}$$

Suppose such limit exist:

$$\zeta \to n\zeta$$

$$\Lambda(\zeta) = \lim_{n\to\infty}\frac{1}{n}\ln\langle e^{n\zeta S_n}\rangle = \lim_{n\to\infty}\frac{1}{n}\ln\prod_{i=1}^N\langle e^{\zeta X_i}\rangle = \ln\langle e^{kX}\rangle \tag{1.29}$$

We have such exponential decay with the greatest contribution from saddle point if $\Lambda(\zeta)$ is analytic:

$$P(S_n = s) \sim \int_{\zeta^*-i\infty}^{\zeta^*+i\infty} d\zeta e^{-n(\zeta s - \Lambda(\zeta))} \sim e^{-n(\zeta^* s - \Lambda(\zeta^*))} \tag{1.30}$$

Which should be the minimum of the $-n(\zeta s - \Lambda(\zeta))$ or the maximum of $\zeta s - \Lambda(\zeta)$.

For Gaussian sample, which is already what we expected in exponential expression, we could evaluate the $\Lambda(k)$:

$$\langle e^{kX}\rangle = \int_{-\infty}^\infty e^{kx}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \tag{1.31}$$

$$-\frac{1}{2\sigma^2}(x^2 - 2(\mu + k\sigma^2)x + \mu^2) = -\frac{1}{2\sigma^2}\left[(x - (\mu + k\sigma^2))^2 - (\mu + k\sigma^2) + \mu^2\right]$$

$$-\frac{1}{2\sigma^2}\left((\mu + k\sigma^2)^2 - \mu^2\right) = k\mu + \frac{1}{2}k^2\sigma^2 \tag{1.32}$$

Thus, the quadratic part is also a Gaussian distribution and then we only have the constant part which reduce to:

$$\Lambda(k) = k\mu + \frac{1}{2}k^2\sigma^2$$

$$I(s) = \sup_k(ks - \Lambda(k)) = \sup_k\left(k(s - \mu) - \frac{1}{2}k\sigma^2\right) = \frac{(s - \mu)^2}{2\sigma^2} \tag{1.33}$$

Which is indeed same as original Gaussian sample.

Few property is worth to be noticed:

$$\Lambda'(0) = \lim_{n\to\infty} \frac{\langle S_n e^{nkS_n}\rangle}{\langle S_n e^{nkS_n}\rangle}\big|_{k=0} = \lim_{n\to\infty}\langle S_n\rangle \tag{1.34}$$

$$\Lambda''(0) = \lim_{n\to\infty} n(\langle S_n^2\rangle - \langle S_n\rangle^2) = \lim_{n\to\infty} n \operatorname{Var}(S_n) \tag{1.35}$$

With Holder inequality:

$$\int d\mu fg \leq \left(\int d\mu f^{\frac{1}{p}}\right)^p \left(\int d\mu g^{\frac{1}{q}}\right)^q \quad \left(\frac{1}{p} + \frac{1}{q} = 1\right) \tag{1.36}$$

$$f \to \mathbb{E}X = \mathbb{E}e^{pk_1 X}, g \to \mathbb{E}Y = e^{qk_2 X}$$

$$\ln(\mathbb{E}(XY)) \leq \ln\left(\left(\mathbb{E}X^{\frac{1}{p}}\right)^p \left(\mathbb{E}Y^{\frac{1}{p}}\right)^p\right)$$

$$\Lambda\left(\frac{1}{p}k_1 + \frac{1}{q}k_2\right) = \frac{1}{p}\Lambda(k_1) + \frac{1}{q}\Lambda(k_2) \tag{1.37}$$

We prove convexity of Λ.

Finally, we have a rather lucky news:

$$\lambda(0) = -I(s = 0)|_{k=0} \stackrel{s \text{ is constant}}{=} \sup_s(-I(s)) = -\inf_s I(s) = 0 \tag{1.38}$$

Rather, we also has the convexity of I:

$$I'(s) = k \quad \Lambda'(k) = s$$

$$I''(s) = \frac{dk}{ds} \quad \Lambda''(k) = \frac{ds}{dk}$$

$$I''(s) = \frac{1}{\Lambda''(k)} > 0 \tag{1.39}$$

However, such won't exist if we have no analyticity of function, like mix of discrete and continuous variable. For example extend a mixed Gaussian, use our knowledge on Gaussian sample mean:

$$P(Y = -1) = P(Y = 1) = \frac{1}{2}$$

$$S_n = \frac{1}{n}\sum_{i=1}^{n} X_i + Y$$

$$P(S_n \in ds | Y = \pm 1) \sim e^{-nI_{\pm}(s)}ds$$

$$I_{\pm} = \frac{(s \pm 1)^2}{2}$$

(1.40)

$$P(S_n \in ds) \sim \frac{1}{2}e^{-nI_+(s)} + \frac{1}{2}e^{-nI_-(s)}ds$$

$$I(s) = \min(I_-(s), I_+(s)) = \begin{cases} I_-(s) & s > 0 \\ I_+(s) & s < 0 \end{cases}$$

(1.41)

Which is a mirror function has both $-1, 1$ as its minimum.

If $I(s)$ has a unique global minimum that $I'(s^*) = k = 0$, then: $s^* = \Lambda'(0) = \lim_{n \to \infty} \langle S_n \rangle$. It indicates such such value won't vanish in the exponential decay and others will concentrate upon, called typical value. Thus we can expand it in quadratic terms: $I(s) \approx \frac{1}{2}I''(s^*)(s - s^*)^2$:

$$P(S_n \in ds) \approx e^{-\frac{n}{2}I''(s^*)(s-s^*)^2}ds \quad \left(I''(s^*) = \frac{1}{\Lambda''(0)} = \frac{1}{\sigma^2}\right)$$

(1.42)

Called *Central Limit Theorem*. Which is a Gaussian sample form in limit.

Now we apply the routine to Binary sample:

$$P(X_i = -1) = P(X_i = 1) = \frac{1}{2}$$

$$\langle e^{kX} \rangle = \cosh k$$

$$\Lambda(k) = \ln \cosh k$$

(1.43)

$$I(s) = \sup_k(ks - \Lambda(k))$$

(1.44)

$$\frac{d(ks - \ln \cosh k)}{dk} = s - \tanh k = 0$$

$$s = \frac{e^k - e^{-k}}{e^k + e^{-k}} \to k = \frac{1}{2}\ln\left(\frac{1+s}{1-s}\right)$$

$$1 - \tanh^2 k = \frac{1}{\cosh^2 k} \to \cosh k = \frac{1}{\sqrt{1 - \tanh^2 k}}$$

$$I(s) = s\frac{1}{2}\ln\left(\frac{1+s}{1-s}\right) - \ln\left(\frac{1}{\sqrt{1-s^2}}\right) = \frac{1}{2}((1+s)\ln(1+s) + (1-s)\ln(1-s)) \quad s \in [-1, 1]$$

(1.45)

We can extend the scalar random variable into the vector valued. Given a sequence $\boldsymbol{\omega} = (\omega_1, ..., \omega_n)$, taking the value on state set Ω, called empirical vector, which is a series of experience output, evaluate as:

$$L(\boldsymbol{\omega})(\mu) = \frac{1}{n} \sum_{i=1}^{n} \delta(\omega_i - \mu) \quad \mu \in \Omega$$

$$k \cdot L = \int_{\Omega} k(\mu) L(\omega)(\mu) d\mu$$

(1.46)

$$\int L(\boldsymbol{\omega})(\mu) d\mu = \frac{1}{n} \sum_{i=1}^{n} \int d\mu \, \delta(\omega_i - \mu) = 1_{\mathcal{B}}$$

(1.47)

Is a probability density on state set. We can calculate such large deviation, extend from scalar to vector dot product.

$$\begin{aligned}
\Lambda(k) &= \lim_{n \to \infty} \frac{1}{n} \ln \langle e^{nk \cdot L} \rangle \\
&= \ln \langle e^{k \cdot L_i} \rangle \\
&= \ln \int_{\Omega} d\mu \, p(\mu) \, e^{k \cdot L_i} \quad (k \cdot L_i = k(\mu)(\omega_i)) \\
&= \ln \int_{\Omega} d\mu \, p(\mu) \, e^{k(\mu)(\omega_i)}
\end{aligned}$$

(1.48)

$$\begin{aligned}
I(s) &= \sup_k (k \cdot s - \Lambda(k)) \\
\frac{\delta \Lambda^*(k)}{\delta k} &= s(\mu) - \frac{p(\mu) e^{k(\mu)}}{\int_{\Omega} d\mu p(\mu) e^{k(\mu)(\omega_i)}} = 0 \\
k(\mu) &= \ln \left(s(\mu) \frac{\int(...)}{p(\mu)} \right) = \ln \left(\frac{s(\mu)}{p(\mu)} \right) + \Lambda(k) \\
I(s) &= \int d\mu s(\mu) \ln \frac{s(\mu)}{p(\mu)} + \int d\mu p(\mu) \Lambda(k) - \Lambda(k) = \int d\mu s(\mu) \ln \frac{s(\mu)}{p(\mu)} = -H(s\|p)
\end{aligned}$$

(1.49)

It suggests that $P(L(\boldsymbol{\omega})(\mu) \sim s(\mu)) \sim e^{-nI(s)}$, which is the cost of our observation on the underneath real probability density $p(\mu)$, called **relative entropy**. After enough observation, we conclude that the observation data should converge to the real underneath probability density. If the support of $s(\mu)$ is larger than $p(\mu)$, the functional will explode, and indicating that the thing never gonna happen as a zero probability.

The relative entropy is our establishment on *general maximal entropy principle*, which is the typical point or equilibrium point around behavior deeply related to maximal entropy.

Previously, we deduce Binary sample in equal probability case, we now choose an arbitrary one:

$$P(X = 1) = p, P(X = -1) = q$$

$$s(X = 1) = p', s(X = -1) = q' \text{ where } s \text{ is Binary with } q' \geq q$$

$$H(s\|p) = q' \ln \frac{q'}{q} + p' \ln \frac{p'}{p} = q' \ln \frac{q'}{q} + (1 - q') \ln \frac{1 - q'}{1 - q}$$

(1.50)

Now suppose we shift from distribution s into expectation value s which is a abuse of notation, but I will clarify that:

$$Xs(X) = s'$$

$$1 - 2q' = s'$$

$$q' = \frac{1 - s'}{2}$$

(1.51)

$$I(s') = H(s\|p)(s') = \frac{1 - s'}{2} \ln \frac{1 - s'}{2} q + \frac{1 + s'}{2} \ln \frac{1 + s'}{2(1 - q)}$$

$$= \frac{1}{2} \left((1 + s) \ln \frac{1 + s}{2p} + (1 - s) \ln \frac{1 + s}{2(1 - q)} \right)$$

(1.52)

Which is same as previous example if we choose $p = \frac{1}{2}$.

Reside, we give another proof of large deviation.

$$\mathbb{E}_{P_{\lambda^*}}(X) = \int X dP_{\lambda^*}(X)$$

$$= \int X e^{\lambda^*(X - \Lambda(\lambda^*, x))} dP(X)$$

$$= \frac{M'(\lambda^*)}{M(\lambda^*)} \quad (M(\lambda^*) = \mathbb{E}(\exp(\lambda^* X)))$$

$$= \Lambda'(\lambda^*) = s \quad \text{(choose an aribitrary one } \lambda^* \text{ to achieve it)}$$

(1.53)

$$P(S_n \subseteq B(s, \delta)) = \int_{B(x,\delta)} dP$$

$$= \int_{B(x,\delta)} \frac{dP}{dP_{\lambda^*}} dP_{\lambda^*}$$

$$= \int_{B(x,\delta)} e^{-n\lambda^*(S_n + \Lambda(\lambda^*, x))} dP_{\lambda^*}$$

$$= e^{-n(\lambda^* x - \Lambda(\lambda^*))} \int_{B(x,\delta)} e^{-n\lambda^*(S_n - x)} dP_{\lambda^*}$$

$$= e^{-n(\lambda^* x - \Lambda(\lambda^*))} P_{\lambda^*}(S_n \subseteq B(s, \delta))$$

(1.54)

If such point s is the concentration point of P_{λ^*}, the probability around such point converge to unit in limit.

$$\text{LHS} \sim \lim_{n \to \infty} e^{-n(\lambda^* x - \Lambda(\lambda^*))}$$

(1.55)

Thus such rate function $I(x) = \lambda x - \Lambda(\lambda)$ should exist at least in a boundary condition. Even though the actual function may not exist. Then such minimum if exist, must be $\sup\{\Lambda^*(x)\}|_\lambda$ with a specified lambda, if we are lucky, it should be the convergence too. We can directly put such as the definition:

$$\lim_{n\to\infty} \frac{1}{n} \ln P(S_n \in X) \sim I_X \tag{1.56}$$

Generally, one could apply to for various function $f(s)$, which is the general Laplacian transformation on function $f(s)$ as $\Lambda(f)$.

$$\Lambda(f) = \lim_{n\to\infty} \frac{1}{a_n} \mathbb{E}\left(e^{a_n f(S_n)}\right) = \sup_s (f(s) - I(s)) \tag{1.57}$$

2 Hamiltonian

Given a phase space with coordinates $\mu = (\boldsymbol{q}, \boldsymbol{p})$ where $\boldsymbol{q} = (q^1, ..., q^N)$ and $\boldsymbol{p} = (p^1, ..., p^N)$. Given a volume in such phase space, denotes the initial condition as $\mathcal{R}(0) = \{\mu(0) \mid \mu(0) \in \mathcal{R}_0\} \overset{\text{time evolution}}{\to} \mathcal{R}(t) = \{\mu(t) \mid \mu(0) \in \mathcal{R}\}$. For a instance time $t \to t + \varepsilon$, we has:

$$
\begin{aligned}
d\mu_{t+\varepsilon} &= \det\left(\frac{\partial \mu_{t+\varepsilon}}{\partial \mu_t}\right) d\mu_t \\
&= \det\left(\delta_{ij} + \varepsilon \frac{\partial \dot{\mu}_{ti}}{\partial \mu_{tj}} + O(t^2)\right) d\mu_t \\
&= \left(1 + \varepsilon \mathrm{Tr}\left(\frac{\partial \dot{\mu}_t}{\partial \mu_t}\right)_i + O(t^2)\right) d\mu_t \quad (\log \det M = \mathrm{Tr} \log M)
\end{aligned}
\tag{2.58}
$$

We haven′t introduce the dynamic equation yet, if it satisfy the canonical form:

$$\mathrm{Tr}(...) = \left(\frac{\partial \dot{q}_{ti}}{\partial q_{ti}} + (q \to p)\right) = \left(\frac{\partial}{\partial q_{ti}} \frac{\partial H}{\partial p_{ti}} - \frac{\partial}{\partial p_{ti}} \frac{\partial H}{\partial q_{ti}}\right) = 0 \tag{2.59}$$

Which is just a reframe of the basic Hamiltonian equation, because we just integrate phase point solely.

That′s, given a distribution $\rho(\mu, t)$:

$$
\begin{aligned}
\rho(\mu, t) &= \rho(\boldsymbol{p}, \boldsymbol{q}, t) \\
\frac{d\rho}{dt} &= \frac{\partial \rho}{\partial t} + \{\rho, H\} = 0
\end{aligned}
\tag{2.60}
$$

Above called *Liouville′s equation*.

A famous theorem is *Poincaré recurrence theorem*, which states that for a system with finite energy, the system will return to a state arbitrarily close to the initial state after a sufficiently long time. This is a consequence of the conservation of phase space volume and the deterministic nature of Hamiltonian dynamics.

Given a distribution $\delta(E - H(\boldsymbol{q}, \boldsymbol{p}))$, one has the possible phase space volume as the total volume T the initial condition can transverse, for example $\frac{p^2}{2m} + \frac{1}{2}m\omega^2 x^2 = H(x, p) = E$. Then we can apply a continuous transformation to the phase space volume, due to **finite energy/volume**, the union space. We clarify such transformation as g_τ with given τ as time parameter or any parameter you want, then we has:

$$\text{T} = \bigcup_{k=0}^{\infty} g_\tau^k \mathcal{R}_0 \quad \mathcal{R}_0 \text{ is initial condition phase volume} \tag{2.61}$$

Thus we must have intersection of possible transverse like k to $k + l$ finitely, otherwise T would be infinite due to infinite disjoint sets.

$$g_\tau^k \mathcal{R}_0 \cap g_\tau^{k+l} \mathcal{R}_0 \neq \emptyset \overset{g \text{ is reversible}}{\rightarrow} \mathcal{R}_0 \cap g_\tau^l \mathcal{R}_0 \neq \emptyset \tag{2.62}$$

Which prove the theorem. It said that for a non-dissipative or closed(**invertible + volume preservation**), **finite** system, the system will emerge recurrence. However, this theorem is not collide with macro equilibrium.
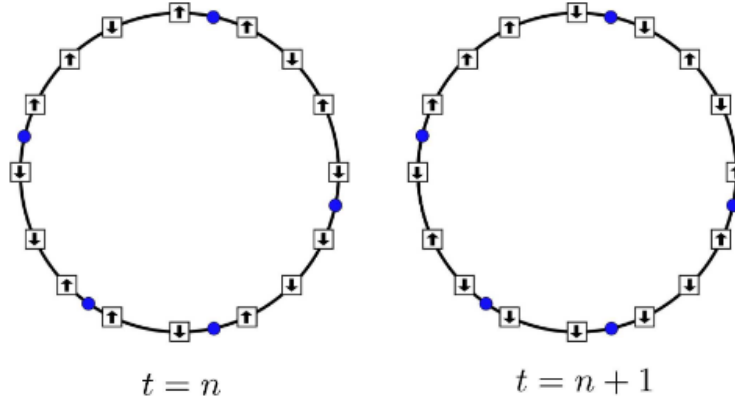
2.1 Kac Ring Model

Figure 1. Left: A configuration of the Kac ring with $N = 16$ sites and $F = 4$ flippers. The flippers, which live on the links, are represented by blue dots. Right: The ring system after one time step. Evolution proceeds by clockwise rotation. Spins passing through flippers are flipped. "

Above picture shows a simple model, N sites is a instance with spin up and down, then we also have F flippers on the link between sites, which can flip spin instance. The evolution is just rotate the ring clockwise one step and the spin instance would be flipped if it pass through the flipper. Thus, it's deterministic and invertible with finite state. It must has **recurrence**.[1]

Now suppose $s(0) = (s_i(0))$ is the initial state, we has a general probability relation:

---

[1]You may wonder why we can apply recurrence to a system without canonical Hamiltonian equation. The reason is that the theorem is not about Hamiltonian system, but about the finite state system with invertible and state volume preservation.

$$s_i(+) = 1, s_i(-) = -1$$

$$\sum_i s_i(+)(n+1) = \frac{N-F}{N} \sum_i s_i(+)(n) + \frac{F}{N} \sum_i s_i(-)(n) \qquad (2.63)$$

$$N^+ + N^- = N$$

$$p^+(n+1) = (1-\eta)p^+(n) + \eta(1-p^+(n)) \quad \eta = \frac{F}{N} \qquad (2.64)$$

$$p^+(n+1) = (1-2\eta)p^+(n) + \eta$$

Thus one should have:

$$p^+(n) = a + b(1-2\eta)^n$$

$$p^+(0) = a + b$$

$$p^+(\infty) = \frac{1}{2} = a \qquad (2.65)$$

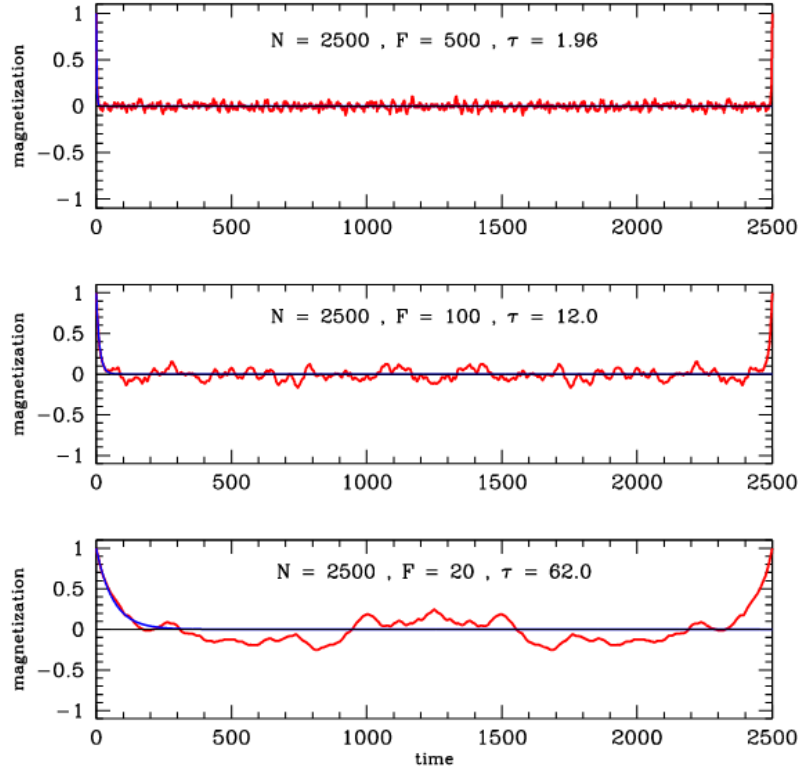$$p^+(n) = \frac{1}{2} + \left(p^+(0) - \frac{1}{2} \right)(1-2\eta)^n$$

Figure 2. Three simulations of the Kac ring model with $N = 2500$ sites and three different concentrations of flippers. The red line shows the magnetization as a function of time, starting from an initial configuration in which 100% of the spins are up. The blue line shows the prediction of the Stosszahlansatz, which yields an exponentially decaying magnetization with time constant τ.

It′s astonishing that given a priori uniform distribution, one can acquire a probability evolution, without any confliction with deterministic evolution, which we cast off information for a specific ring and presume a potential microstates in priori uniform distribution.

Now we construct the microscopic state:

$$\boldsymbol{X}(t) = (s_1(t), s_2(t), ..., s_N(t) | \; m_1(t), m_2(t), ...m_F(t)) \quad (s_{i+1}(t) = m_i s_i(t)) \tag{2.66}$$

For each micro state, one can generate by permutation choices of flippers, which is $\binom{N}{F}$, however we know many of them is same for unknown observer, such process called **coarse-graining** or **projection**, which is often the averaging over irrelevant degrees of freedom, losing micro information increasing macro entropy.

Now the macro state is evident that:

$$\Delta(t) = \sum_i s_i(t) \quad \left(s_i(+) = 1, s_{i(-)} = -1\right) \tag{2.67}$$

With such projection:

$$\Delta(t) = \sum_{\boldsymbol{m}} P(\boldsymbol{m}) \Delta(t|\boldsymbol{m}) \tag{2.68}$$

For each general ensemble of flipper distribution, which is priori uniform distribution, one can decompose as:

$$
\begin{aligned}
\Delta(t) &= \sum_{\boldsymbol{m}} \frac{1}{\binom{N}{F}} \Delta(t|\boldsymbol{m}) \\
&= \sum_{\boldsymbol{m}} \frac{1}{\binom{N}{F}} \sum_i s_i^+(t|\boldsymbol{m})
\end{aligned}
\tag{2.69}
$$

$$\sum_i s_i^+(t|\boldsymbol{m}) = \sum_i \left(\prod_{k=0}^{t-1} m_{i-k} s_i(0) \right) = \prod_{k=0}^{t-1} m_{i-k} \Delta(0) \tag{2.70}$$

The key is the uniform distribution that:

$$\langle \prod_{k=0}^{t-1} m_{i-k} \rangle = \prod_{k=0}^{t-1} \langle m_{i-k} \rangle = \prod_{k=0}^{t-1} (-P(m_i = -1) + P(m_i = +1)) = \prod_{k=0}^{t-1} (1 - 2\eta) = (1 - 2\eta)^t \tag{2.71}$$

Thus, the distribution of $\boldsymbol{m}$ is actually unimportant, which is cancel out.

$$\Delta(t) = (1 - 2\eta)^t \Delta(0) \tag{2.72}$$

$$
\begin{aligned}
\Delta^+(t) + \Delta^-(t) &= \Delta(t) \\
\Delta^+(t) - \Delta^-(t) &= N
\end{aligned}
\tag{2.73}
$$

$$
\begin{aligned}
\Delta^+(t) &= \frac{1}{2}(N + \Delta(t)) = \frac{1}{2} \sum_i (1 + s_i(t)) \\
\Delta^+(t) &= \frac{1}{2}(N + (1 - 2\eta)^t \Delta(0))
\end{aligned}
\tag{2.74}
$$

Which is indeed the original solution. That′s more rigorous process demonstrate how we establish the distribution in blurring information, in which we lose initial condition and presume flippers counts only. Following from the probability investigation, we can therefore recover the **true** probability distribution based on the test, because there′s no guarantee a uniform distribution for potential flippers distribution.

If we consider the all micro state entropy, it should be fixed that transverse all **priori** uniform distribution. rather, if we consider an **ensemble** of rings in macro state, one has:

$$H_{\mathrm{micro}}(\mathrm{up/down}) = \ln \Omega = \ln 2^N = N \ln 2 \quad \text{fixed} \tag{2.75}$$

$$H(\text{flippers}) = \ln\left(\binom{N}{M}\right) = \ln\left(\frac{N!}{M!(N-M)!}\right)$$

$$\sim N\ln N - N - M\ln M + M - (N-M)\ln(N-M) + (N-M)$$

$$\sim -N\left(\frac{M}{N}\ln\frac{M}{N} - \left(1 - \frac{M}{N}\right)\ln\frac{N-M}{N}\right)$$

$$\sim -N(\eta\ln\eta + (1-\eta)\ln(1-\eta)) \quad \text{fixed}$$

(2.76)

$$H_{\text{macro}}(\text{up/down})(t) = -N(p\ln p + (1-p)\ln(1-p)) \quad p(t) = \frac{\Delta^+(t)}{N}$$

(2.77)

It has the maximum in $p = \frac{1}{2}$ as the convergence results, showing the entropy increasing dynamics based on macro state, reaching the micro state entropy.

Now we can summarize to answer the problem of *Hamilton Dynamics*, why micro determinism will emerge macro uncertainty? It's plausible to say information is a key, first, it's objective but relative, like speed, it's a coarse-graining process to shed underneath initial condition, we have no idea about the concrete position and momentum but a **priori** potential distribution of them, from which, the observation tells us, in a enough time, one should approach the real distribution.

3 Ergodicity

A stronger statement is *ergodicity*, it means given all possible phase volume T, the distribution of phase point in a longer enough time would transverse all possible phase point in T, in $t \to \infty$:

$$\langle \rho(\mu) \rangle_t = \lim_{T\to\infty}\left(\frac{1}{T}\right)\int_0^T d\mu\rho(\mu t)$$

$$= \int d\mu\rho(\mu)\delta(E - H(\mu))/\int d\mu\delta(E - H(\mu))$$

$$= \langle \rho(\mu) \rangle_E$$

(3.78)

Usually, denoted as the hyper-surface as:

$$D(E) = \int d\mu\delta(E - H(\mu)) = \int_{S_E} d\Sigma_E$$

(3.79)

The integral inspection doesn't ensure the stability of the distribution, clearly, we can have a distribution with oscillation in time and with a suitable boundary condition, which make cause equality, but the oscillation property of the distribution won't disappear in time.

$$f(q,p) = \sum_{m,n}\hat{f}_{m,n}e^{imq+inp}$$

$$\langle f(q,p)\rangle_t = \hat{f}_{0,0} = \langle f(q,p)\rangle_E$$

(3.80)

4 Fundamental

For particle n in system with particles N in fixed energy bound that $H \in dU$ as a macro state in a possible range or $H = U_0$ if you want, which can be translated to $h_n \in du$, but the latter should be preferred in **thermodynamic limit**, also called **large-number limit**.

$$P(h_n \in du) = \int_{\{h_n(\omega) \in du | \omega \in \Omega_n\}} P(d\omega) \tag{4.81}$$

$$\Omega(h_n) = \int_{\{h_n(\omega) \in du | \omega \in \Omega_n\}} d\omega \tag{4.82}$$

For uniform prior measure: $P(d\omega) = \frac{d\omega}{|\Omega_n|^N}$

$$I(u) = \lim_{n \to \infty} -\frac{1}{n} \ln P(h_n \in du) = \ln|\Omega_n| - I'(u)$$

$$I'(u) = \lim_{n \to \infty} \frac{1}{n} \ln \Omega_n(h_n \in du) \tag{4.83}$$

That's, given the energy bound as $\delta(E - H(\mu))$, one has a energy shell that contains the possible states. In a uniform prior measure we can reduce the constant and use the latter $I'(u)$, which called $s(u)$ in classical expression.

$$\lambda(\beta) = \lim_{n \to \infty} \frac{1}{n} \ln \langle e^{n\beta h_n} \rangle = -\lambda'(-\beta) - \ln|\Omega_n|$$

$$\lambda'(\beta) = \lim_{n \to \infty} -\frac{1}{n} \ln \int_{\Omega_n} e^{-n\beta h_n(\omega)} d\omega = \lim_{n \to \infty} -\frac{1}{n} \int_{\Omega_n} e^{-\beta H(\omega)} d\omega \tag{4.84}$$

Where the latter function is the partition function associated with $H(\omega)$, is the canonical free energy function, denoted as $\lambda'(\beta) = \varphi(\beta)$ too with a intended constant reduction and variable transformation.

We have above *Legendre-Fenchel Transformation* which is the saddle point transformation of *Laplace Transformation*.

$$\varphi(\beta) = \inf_u (\beta u - s(u))$$

$$s(u) = \inf_\beta (\beta u - \varphi(\beta)) \tag{4.85}$$

Due to the common notation about negative sign, from $\sup \to \inf$ is evident. If the whole is differentiable, it should be convex and contains global minimum, else, it contains multiple, or **phase transformation** in such non-differentiable configuration.

Suppose now, given a macro state, the constrained probability measure is given by:

$$P^M(M_n \in dm) = P^M(M_n \in dm | h_n \in du) = \frac{P(h_n \in du, M_n \in dm)}{P(h_n \in du)} \tag{4.86}$$

With the constraint entropy based on the macro state, and the energy represented by the macro state:

16

$$P^M(M_n \in dm) \sim e^{-nI^M(u)} du$$

$$I^M(u) = s(u) - s^M(u) \quad \text{if } h(M_n(\omega)) = u \tag{4.87}$$

Where $s^M(u)$ is the entropy of joint probability. Denoting the typical values or equilibrium values(states):

$$\mathcal{E}^M = \{m : I^M(m) = 0\} = \{m : m \text{ globally maximize } s^M(m)\} \tag{4.88}$$

Which the stable point is $s(u) = \sup_m s^M(u, m)$ or $\sup_{\{m|h^{M(m)}=u\}} s^M(m)$, that can be used to calculate the micro canonical entropy from the knowledge of a macro state entropy[2] as the **fundamental of equilibrium statistical mechanics**.

Macrostate:

$$L_n(\varepsilon) = \frac{1}{n} \sum_{i=1}^{n} \delta(\varepsilon_i - \varepsilon) \tag{4.89}$$

From Sanov′ s theorem, the large deviation of L_n are ruled by the relative entropy $I^M(\mu)$ given the prior uniform measure.

Energy representation and normalization as constraint:

$$1 = \int p(\varepsilon) d\varepsilon$$

$$h^\varepsilon(p) = \int p(\varepsilon) \varepsilon d\varepsilon \tag{4.90}$$

Macrostate entropy:

$$s^M(p^M) = \int d\varepsilon \, p^M(\varepsilon) \ln p^M(\varepsilon) \tag{4.91}$$

We maximize by Lagrange multiplier:

$$S\{p^M, h^M\} = s^M(p^M) + \lambda_1 (h^M(p^M) - u) + \lambda_2 \left(\int p^M(\varepsilon) d\varepsilon - 1 \right) \tag{4.92}$$

$$p^M(\varepsilon) = \frac{e^{-\beta\varepsilon}}{Z(\beta)}$$

$$Z(\beta) = \int d\varepsilon \, e^{-\beta\varepsilon} \quad (\lambda_1 = -\beta) \tag{4.93}$$

Therefore, one has the maximized macro entropy as micro entropy:

$$s(p) = s^M(p) - \ln|\Omega| = \beta\varepsilon - \varphi(\beta) \tag{4.94}$$

---

[2]You already have seen this in Kac Ring model.

Where $\varphi(\beta)$ is the free energy.

Reversely, we can control the parameter in the other side of Legendre-Fenchel transformation, which is β. Given the free energy configuration defined previously:

$$I^{*M}(m) = s(u) - s^M(u) \sim \varphi^M(\beta) - \varphi(\beta)$$
$$\mathcal{E}^{*M} = \{m|I^{*M}(m) = 0\} = \{m|m \text{ is global minimum of } I^{*M}(m)\}$$
(4.95)

Thus you may want to minimize the free energy which is $\beta h^M(m) - s^M(m)$. That's:

$$S^*\{p^M, h^M\} = \beta h^M - s^M - \lambda\left(\int d\varepsilon p(\varepsilon) - 1\right)$$
(4.96)

$$p^M(\varepsilon) = \frac{e^{-\beta\varepsilon}}{Z(\beta)} p(\varepsilon)$$
$$I^M(\varepsilon) = \beta u - \varphi(\beta) - s(u) = s^M(u) - s(u) = s^M(u) - \ln|\Omega|$$
(4.97)

... The class of macrostates for which large deviation principles can be derived in the canonical ensemble is exactly the same as in the microcanonical ensemble, since the rate functions of these two ensembles are built from the same energy. representation function and macrostate entropies.

— The large deviation approach to statistical mechanics — Hugo Touchette

The second way is more preferred, which the partition functional is easier to calculated and suitable for deduction of many typical parameters.

$$Z(\beta) = \int d\omega\, e^{-\beta H(\omega)} = \int d\omega\, e^{\lambda E}\delta(H(\omega) - E) = \int d\omega\, e^{\lambda E} D(E) = \mathcal{L}(D(E))$$
(4.98)

$$P(\mu) = \frac{1}{Z}\exp\left(\sum_a \lambda_a X^a(\mu)\right)$$
$$\frac{\partial \ln Z}{\partial \lambda_a} = \langle X^a(\mu)\rangle$$
(4.99)

$$S[P(\mu)]|_{\text{extreme}} = \int d\mu P(\mu)\log P(\mu)|_{\text{extreme}}$$
$$= \int d\mu P(\mu)(\lambda H(\mu) - \log Z)$$
$$= \lambda E - \log Z$$
(4.100)

$$\lambda^{-1}\log Z = E - \lambda^{-1}S[P(\mu)]|_{\text{extreme}} \rightsquigarrow F = \lambda^{-1}\log Z = E - TS$$
(4.101)

Given a exchange ensemble with particle number N and volume V, we has:

$$P(\mu, N) = \frac{1}{Z} \exp(\lambda_1 H(\mu) + \lambda_2 N(\mu))$$

$$\frac{\partial \ln Z}{\partial \lambda_1} = E \tag{4.102}$$

$$\frac{\partial \ln Z}{\partial \lambda_2} = \langle N(\mu) \rangle = N$$

$$\lambda_1 = \lambda_{1'} \text{ in equalibrium}$$
$$\lambda_2 = \lambda_{2'} \text{ in equalibrium} \tag{4.103}$$
$$\rightsquigarrow \frac{\lambda_1}{\lambda_2} = \frac{\lambda_{1'}}{\lambda_{2'}}$$

$$S[P(\mu)]|_{\text{extreme}} = \lambda_1 E + \lambda_2 N - \log Z$$
$$\lambda_1^{-1} \log Z = E + \lambda_2 \lambda_1^{-1} N - \lambda^{-1} S[P(\mu)]|_{\text{extreme}} \rightsquigarrow G = E - uN - TS \tag{4.104}$$

$$\frac{\partial \lambda_1^{-1} \ln Z}{\partial \lambda_2 / \lambda_1} = N \tag{4.105}$$

5 Selected Example

5.1 Ideal Gas

Given general parameters, we have general exchange system or grand canonical system:

$$\lambda_1^{-1} d \log Z = \overline{X_a} d(\lambda_a \lambda_1^{-1})$$
$$d \log Z^* = \overline{X_a} d\lambda_a^* \tag{5.106}$$

$$H(\mu) = \frac{1}{2m} \sum_i^{3N = dN} p_i^2 = \sum_i^{3N = dN} u_i^2 \quad u_i = \sqrt{2mE} p_i \tag{5.107}$$

$$D(E, V, N) = \int \frac{1}{N!(2\pi\hbar)^d} d^{Nd} q\, d^{Nd} p\, \delta\left(\frac{1}{2m} \sum_i^{dN} p_i^2 - E\right)$$
$$= \frac{V^N}{N!} \left(\frac{\sqrt{2mE}}{h}\right)^{Nd} \frac{1}{E} \int d^M u\, \delta\left(\sum_i^M u_i^2 - 1\right) \quad (M = Nd)/(\delta(Ex) = E^{-1}\delta(x)) \tag{5.108}$$

Where we integral on a solid ball with unit radius, thus the radial part integral must be unit. Also, the delta function will be reduced to $\delta\left(\sum_i u_i^2 - 1\right) = \delta(\boldsymbol{u}^2 - 1) = \frac{1}{2}(\delta(\boldsymbol{u} + 1) + \delta(\boldsymbol{u} - 1))$ with ball integral $d^M u = \boldsymbol{u}^{M-1} d\boldsymbol{u} d\Omega_M$.

$$D(E, V, N) = \frac{V^N}{N!} \left(\frac{\sqrt{2m}}{h}\right)^{Nd} E^{\frac{1}{2}Nd - 1} \frac{1}{2} \Omega_{Nd} \tag{5.109}$$

$$\Omega_M = \frac{2\pi^{\frac{M}{2}}}{\Gamma\left(\frac{M}{2}\right)} \tag{5.110}$$

$$Z(T, V, N) = \int \frac{1}{N! 2\pi\hbar^d} d^{Nd}q \, d^{Nd}p \, e^{-\beta \boldsymbol{p_i}^2/(2m)}$$

$$= \frac{V^N}{N!} \left(\int_{-\infty}^{\infty} \frac{dp}{2\pi\hbar} e^{-\beta p^2/(2m)} \right)^{Nd} \tag{5.111}$$

$$= \frac{V^N}{N!} \left(\frac{1}{h} \sqrt{\frac{2\pi m}{\beta}} \right)^{Nd} \sim \left(\sqrt{\frac{m}{2\pi\hbar\beta}} \right)^{Nd} = \frac{V^N}{N!} \left(\frac{1}{\lambda_T} \right)^{Nd}$$

$$\Xi(T, V, \mu) = \sum_{N=0}^{\infty} e^{\beta\mu N} Z(T, V, N) = \sum_{N=1}^{\infty} \frac{1}{N!} \left(\frac{V e^{\beta\mu}}{\lambda_T^d} \right)^N = \exp\left(\frac{V e^{\beta\mu}}{\lambda_T^d} \right) \tag{5.112}$$

Now we calculate basic quantity:

$$\log Z^* = \overline{X}_a d\lambda_a^*$$

$$\beta^{-1} \log Z(T, V, N) = \beta^{-1} (N \log V - Nd \log \lambda_T - \log N!) \tag{5.113}$$

$$P_c = -\frac{\partial \log Z^*}{\partial V} = \frac{N\beta^{-1}}{V} = \frac{Nk_bT}{V} \tag{5.114}$$

$$E = -\frac{\partial \log Z}{\partial \beta} = \frac{Nd}{2} \beta^{-1} = \frac{Nd}{2} k_b T \tag{5.115}$$

The notorious *Maxwell Velocity Distriution* can also be established by given empirical vector of momentum.

$$L(\boldsymbol{p}) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \delta(p_i - p) \tag{5.116}$$

$$\langle L(\boldsymbol{p}) \rangle = \frac{\int d^d p_1 \delta(p_1 - p) e^{-\beta p_1^2/2m}}{\int d^d p_1 e^{-\beta p_1^2/2m}} \tag{5.117}$$

The key here is the probability distribution for each particle is already coded as the thermodynamics limit of them, which are same in limit, that $\frac{1}{n} \cdot n \int (...) = \int (...)$, second, else factors are same thus are cancelled out. Finally we leaving:

$$\text{LHS} = \left(\frac{\beta}{2\pi m} \right)^{\frac{d}{2}} e^{-\beta p^2/2m} \tag{5.118}$$

Sometimes, one will transform the variables to $d^3p \to p^2 dp d\Omega = 4\pi p^2 dp$ from $\boldsymbol{p} \to |p|$ relation.

5.2 Spins and External Magnetic Field

Use hat to denote the Hamiltonian else is the magnetic field tensity.

$$\sigma(\text{up}) = 1, \sigma(\text{down}) = -1$$

$$\hat{H} = -\mu_0 H \sum_{i=1}^{N} \sigma_i \tag{5.119}$$

We should transverse all states of σ for each i:

$$Z = \sum_{\sigma_1} \cdots \sum_{\sigma_N} e^{-\beta \hat{H}} = \sum_{\sigma_1} e^{\beta \mu_0 H \sigma_1} \cdots \sum_{\sigma_N} e^{\beta \mu_0 H \sigma_N} = \prod_{i}^{N} \sum_{\sigma_i} (\ldots) = 2 \cosh^N (\beta \mu_0 H) \tag{5.120}$$

$$F(T, H, N) = -\beta^{-1} \log Z = -N\beta^{-1} \log(2 \cosh(\beta \mu_0 H)) \tag{5.121}$$

$$M = -\frac{\partial F}{\partial H} = N\mu_0 \tanh \beta \mu_0 H$$

$$E = \frac{\partial(\beta F)}{\partial \beta} = -N\mu_0 H \tanh(\beta \mu_0 H) \tag{5.122}$$

$$E = -MH \sim \hat{H} = -\mu_0 H \sum_{i=1}^{N} \sigma_i$$