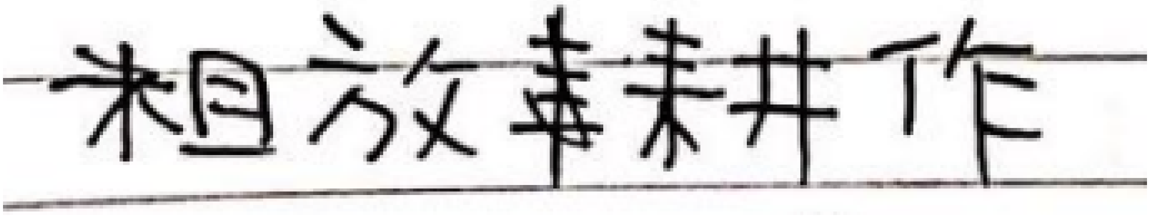

概率论进阶笔记

黄盛唐，Peanut Tang

2024 年 6 月 18 日

目录

1 随机矩阵

1.1 基本模型

Definition 1.1. 在概率空间 $(\Omega, \mathcal{F}, \Pr)$ 里，一个随机变量为 Borel 可测函数 $X : \Omega \to \mathbb{R}$。可以扩展出随机向量 $X : \Omega \to \mathbb{R}^n$、随机矩阵 $X : \Omega \to \mathbb{R}^{n\times n}$。

关心：模型、现象（半圆律）。

参考书：Tao: Topics in Random Matrix Theory. AMS, 2012.

随机矩阵的起源与研究动机。

Example 1.2 (Wishart, 1928)**.** 背景是多元统计中的样本协方差矩阵。

一个 p 维数据 $X_i : \Omega \to \mathbb{R}^{p\times 1}$，我们采样若干次，取平均得到样本协方差矩阵：

$$W = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^T = \frac{1}{n} X X^T$$

这里 $X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^{p\times n}$。

若 $\{X_n\}$ 相互独立，且均服从 $\mathcal{N}(0, I_n)$，则称 W 为一个 Wishart 矩阵。

此时 X 的矩阵元联合密度函数为（自变量为矩阵）：

$$f_X(V) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\operatorname{tr}\left(VV^T\right)\right)$$

Example 1.3 (Wigner, 1955)**.** 背景是量子物理。

我们有 Schrödinger 方程：

$$\frac{\partial}{\partial t}\Psi(x,t) = H\Psi(x,t)$$

这里 $H = -\dfrac{\partial^2}{\partial x^2} + V(x)$。此方程变量多，相互作用复杂，没有额外假设下十分难解。

我们引入随机性，并且将无穷维的算子 H 截断，得到厄米矩阵 $H_n \in \mathbf{U}^n$，其中矩阵元为随机变量。研究 $n \to \infty$ 时 H_n 谱的渐进性，以此来得到 H 算子的一些性质。

下面随机矩阵的一些基本模型。

Definition 1.4 (高斯正交系综)**.** 所谓高斯正交系综（简称 GOE，Gaussian Orthogonal Ensemble）指：

$$A_n = \frac{1}{2}\left(X_n + X_n^T\right)$$

这里 $X_n = (x_{ij})_{i,j=1}^n$，$\{x_{ij}\}$ i.i.d. $\mathcal{N}(0, \sigma^2)$。

我们记 $A_n \sim \mathrm{GOE}(\sigma)$。

可以看出 A_n 上三角部分的矩阵元是相互独立的。我们可以写出每一个矩阵元的分布：

$$a_{ii} = x_{ii} \sim \mathcal{N}(0, \sigma^2)$$

$$a_{ij} = a_{ji} = \frac{1}{2}(x_{ij} + x_{ji}) \sim \mathcal{N}(0, \sigma^2/2)\ (i < j)$$

依此可写出矩阵元的联合密度函数：

$$f_{A_n}(V) = \frac{1}{2^{n/2}\cdot(\pi\sigma^2)^{n(n+1)/4}} \exp\left(-\frac{1}{2\sigma^2}\operatorname{tr}\left(V^2\right)\right)$$

Definition 1.5 (高斯酉系综). 类似的可以定义高斯酉系综（简称 GUE，Gaussian Unitary Ensemble）：

$$A_n = \frac{1}{2}\left(X_n + X_n^*\right)$$

这里 $X_n = (x_{ij})_{i,j=1}^n$，$\{x_{ij}\}$ i.i.d. $\mathcal{N}_{\mathbb{C}}(0,\sigma^2)$。而 $z \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2)$ 为复高斯分布，有概率密度函数：

$$f(z) = \frac{1}{\pi\sigma^2}\exp\left(-\frac{|z|^2}{\sigma^2}\right)$$

我们记 $A_n \sim \mathrm{GUE}(\sigma)$。

可以看出 A_n 上三角部分的矩阵元是相互独立的。我们可以写出每一个矩阵元的分布：

$$a_{ii} = \mathrm{Re}(x_{ii}) \sim \mathcal{N}(0,\sigma^2/2)$$

$$a_{ij} = a_{ji}^* = \frac{1}{2}(x_{ij} + x_{ji}^*) \sim \mathcal{N}_{\mathbb{C}}(0,\sigma^2/2)\ (i<j)$$

依此可写出矩阵元的联合密度函数：

$$f_{A_n}(V) = \frac{2^{n(n-1)/2}}{(\pi\sigma^2)^{n^2/2}}\exp\left(-\frac{1}{2\sigma^2}\operatorname{tr}\left(V^2\right)\right)$$

Theorem 1.6. 我们有以下结论：

(1) GOE 具有正交变换下的不变性，即对任意的正交矩阵 $Q \in \mathrm{O}(n)$ 与 $A_n \sim \mathrm{GOE}(\sigma)$，有 QA_nQ^T 与 A_n 同分布。

(2) 同样的,GUE 具有酉变换下的不变性,即对任意的酉矩阵 $U \in \mathrm{U}(n)$ 与 $A_n \sim \mathrm{GUE}(\sigma)$，有 UA_nU^* 与 A_n 同分布。

Proof. (1) 对于 $X \sim \mathrm{GOE}(\sigma)$，$Q \in \mathrm{O}(n)$，令 $X_i = (X_{i1}, X_{i2}, \ldots, X_{in})$ 为 X 的第 i 行，可以发现有 $X_i \sim \mathcal{N}(0,\sigma^2 I_n)$。

则由正态分布的经典结论：$X_iQ^T \sim \mathcal{N}(0Q^T, Q\sigma^2 I_n Q^T) = \mathcal{N}(0,\sigma^2)$，与 X_i 同分布。故 XQ^T 与 X 同分布。

同理，再左乘一个 Q 也同分布，即 QXQ^T 与 X 同分布。故可以推出 $QXQ^T = \frac{1}{2}(QXQ^T + (QXQ^T)^T)$ 与 X 同分布。

(2) 和 GOE 一样，我们证明 XU^* 与 X 同分布，令 $X_i = (X_{i1}, X_{i2}, \ldots, X_{in})$ 为 X 的第 i 行，可以发现有 $\mathrm{Re}(X_i), \mathrm{Im}(X_i) \sim \mathcal{N}\left(0, \frac{1}{2}\sigma^2 I_n\right)$。

故由实正态分布的经典结论：$\mathrm{Re}(X_iU^*) \sim \mathcal{N}\left(0\,\mathrm{Re}(U^*), \mathrm{Re}(U^*)^T \cdot \frac{1}{2}\sigma^2 I_n \cdot \mathrm{Re}(U^*)\right) + \mathcal{N}\left(0\,\mathrm{Im}(U^*), \mathrm{Im}(U^*)^T \cdot \frac{1}{2}\sigma^2 I_n \cdot \mathrm{Im}(U^*)\right) = \mathcal{N}\left(0, (\mathrm{Re}(U^*)^2 - \mathrm{Im}(U^*)^2) \cdot \frac{1}{2}\sigma^2 I_n\right)$。

因为 $U \in \mathrm{U}(n)$，所以 $UU^* = (\mathrm{Re}(U) + i\,\mathrm{Im}(U))^2 = \mathrm{Re}(U)^2 - \mathrm{Im}(U)^2 = I_n$，所以 $\mathrm{Re}(X_IU^*) \sim \mathcal{N}\left(0, \frac{1}{2}\sigma^2 I_n\right)$。

同理 $\mathrm{Im}(X_iU^*) \sim \mathcal{N}\left(0, \frac{1}{2}\sigma^2 I_n\right)$，故 X_iU^* 与 X_i 同分布，之后同 (1)。

这实际上也证明了复高斯分布中类似的结论——酉变换不变性。 □

1.2 Wigner 半圆律

再来介绍一个随机矩阵里的重要且基本的结果，Wigner 半圆律。

参考文献：Wigner: Annals of Math, 1955.

Lemma 1.7. 定义：

$$\rho_{\mathrm{sc}}(x) = \frac{1}{2\pi}\sqrt{4-x^2},\ x \in [-2,2]$$

可以看出这是一个概率密度函数，我们称为 Wiger 半圆律（分布）。

对比标准正态分布 $X \sim \mathcal{N}(0,1)$，有密度函数 $f(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)$，并且我们可以算出各阶矩：

$$\mathbb{E}[X^k] = \begin{cases} 0 & k \text{ odd} \\ (2m-1)!! & k = 2m \end{cases}$$

对于 $\rho_{\mathrm{sc}}(x)$ 为密度的随机变量，我们也可以算其各阶矩：

$$\int_{-2}^{2} x^k \rho_{\mathrm{sc}}(x)\,\mathrm{d}x = \begin{cases} 0 & k \text{ odd} \\ \dfrac{1}{m+1}\dbinom{2m}{m} & k = 2m \end{cases}$$

可以看到 $k = 2m$ 时对应阶矩为 Catalan 数 C_m。

Proof. 当 k 为奇数是显然对应阶矩为 0。当 $k = 2m$ 时：

$$\begin{aligned}
\int_{-2}^{2} x^k \rho_{\mathrm{sc}}(x) &= \frac{1}{\pi}\int_0^2 x^k\sqrt{4-x^2}\,\mathrm{d}x \\
&= \frac{2^{2m+2}}{\pi}\int_0^{\pi/2} \sin^k\theta\cos^2\theta\,\mathrm{d}\theta\ (x = 2\sin\theta) \\
&= \frac{2^{2m+2}}{\pi}\cdot\frac{1}{2}\mathrm{B}\left(m+\frac{1}{2},\frac{3}{2}\right) = \frac{2^{2m+1}}{\pi}\cdot\frac{\Gamma\left(m+\frac{1}{2}\right)\Gamma\left(\frac{3}{2}\right)}{\Gamma(m+2)} \\
&= \frac{2^{2m+1}}{\pi}\cdot\frac{\frac{(2m-1)!!}{2^m}\sqrt{\pi}\cdot\frac{1}{2}\sqrt{\pi}}{(m+1)!} = \frac{1}{m+1}\binom{2m}{m} = C_m
\end{aligned}$$

得证。 □

Remark. $(2m-1)!!$ 的组合解释：$1, 2, \ldots, 2m$ 两两配对不计顺序方案数。

C_m 的组合解释：$1, 2, \ldots, 2m$ 不自交的两两配对方案数。

不自交：举例，当 $m = 2$ 时，将 $1, 2, 3, 4$ 画在平面上的一条线上，配对的两个数在平面的同一边用一条线连起来。这样 $(1,2),(3,4)$ 与 $(1,4),(2,3)$ 是不自交的，而 $(1,3),(2,4)$ 是自交的。可以用长为 $2m$ 的合法括号串来一一对应。

Definition 1.8 (Wigner 矩阵)**.** 设 $A_n = (a_{ij}) \in \mathbf{S}^n(\mathbb{R})$ 为一实对称随机矩阵，若其矩阵元满足如下要求，则称 A_n 为 Wigner 矩阵：

- $\{a_{ij}\}_{1\le i\le j\le n}$ 相互独立。

- $\{a_{ii}\}$ 与 Y 同分布，而 $\{a_{ij}\}_{i<j}$ 与 Z 同分布。且 Y, Z 满足：$\mathbb{E}[Y] = \mathbb{E}[Z] = 0$，$\mathrm{Var}(Z) = 1$，$\mathrm{Var}(Y) < \infty$，以及 $\forall k \geq 3$, $\mathbb{E}[|Y|^k], \mathbb{E}[|Z|^k] < \infty$。

Definition 1.9 (经验分布)**.** 给定一个实对称（随机）矩阵 $A \in \mathbf{S}^n(\mathbb{R})$ 的特征值 $\lambda_1(A), \ldots, \lambda_n(A)$。令 $F_A(x)$ 为：

$$F_A(x) = \frac{1}{n}\#\{j \mid \lambda_j(A) \leq x\}$$

可以看出 $F_A(x)$ 是一个分布函数，称为 A 谱的经验分布（简称为 ESD，Empirical Spectral Distribution）。

当 A 为随机矩阵时，$F_A(x)$ 为随机 ESD（随机分布函数的具体定义按下不表）。

对于 Borel 可测函数 $g : \mathbb{R} \to \mathbb{R}$（测试函数），可以定义抽象积分：

$$\mathbb{E}[g(A)] = \int g \, \mathrm{d}F_A = \frac{1}{n}\sum_{i=1}^{n} g(\lambda_j) = \frac{1}{n}\operatorname{tr}(g(A))$$

当 A 为随机矩阵时，上面的抽象积分本身也是一个随机变量（两重随机性）。

当 g 为多项式函数时，可以拆分为一个个单项式，实际上就是在考察矩的信息。而矩的信息可以反过来确定分布函数（特征函数）。

Theorem 1.10 (Wigner 半圆律)**.** 对于 Wigner 矩阵 $\{A_n\}$：

$$\frac{1}{n}\,\mathbb{E}\left[\operatorname{tr}\left(\left(\frac{A_n}{\sqrt{n}}\right)^k\right)\right] \to \gamma_k = \begin{cases} 0 & k \text{ odd} \\ C_m & k = 2m \end{cases}$$

由矩收敛定理知：$F_{A_n}(x)$ 弱收敛到 $\rho_{\mathrm{sc}}(x)$。

Remark. 对比 CLT：$\{X_n\}$ i.i.d.，$S_n = \sum\limits_{i=1}^{n} X_i$，$\mathbb{E}[X_i] = 0$，$\mathbb{E}[X_i^2] = 1$，且 X_i 各阶矩一致有界，则：

$$\mathbb{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right] \to \begin{cases} 0 & k \text{ odd} \\ (2m-1)!! & k = 2m \end{cases}$$

同样由矩收敛定理知：$\dfrac{S_n}{\sqrt{n}} \xrightarrow{D} \mathcal{N}(0,1)$。

Proof. 有：

$$LHS = \frac{1}{n^{1+k/2}} \sum_{i_1, i_2, \ldots, i_k = 1}^{n} \mathbb{E}[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_k i_1}]$$

和矩方法证明 CLT 时类似，我们还是对右边的指标进行分类。

把 $i_1, i_2, \ldots, i_k$ 看成顶点，而 $i_1 i_2, \ldots, i_k i_1$ 看成边，这实际上是一个图上的环，从 i_1，走到 i_k，最后再回到 i_1。

只看主项，主项里不能有一条边只出现一次（否则根据独立性以及条件 $\mathbb{E}[Y] = \mathbb{E}[Z] = 0$，会得出这一项为 0，没有贡献），故主项里每条边至少出现两次。

又因为有 $\dfrac{1}{n^{1+k/2}}$ 对主项次数做限制，所以肯定自由度越高越好，我们让不同边的个数达到上界 $\left\lfloor \dfrac{k}{2} \right\rfloor$，此时 $i_1, i_2, \ldots, i_k$ 有 $\left\lfloor \dfrac{k}{2} \right\rfloor + 1$ 个不同顶点。

当 k 为奇数时，$\left\lfloor \frac{k}{2} \right\rfloor + 1 < \frac{k}{2} + 1$，所以没有主项，$n \to \infty$ 时极限为 0。

当 $k = 2m$ 时，考虑计数合法环的个数。发现取定起点后，我们可以和合法括号串一一对应（$i_s \to i_t$ 边为'('，反向边 $i_t \to i_s$ 为')'），所以合法环数量是选取 $m+1$ 个自由元的方案数 $n(n-1)\cdots(n-m)$，乘上 Catalan 数 C_m，$n \to \infty$ 时极限为 C_m。 □

2 中心极限定理重访

2.1 矩收敛定理

在概率论本篇和外篇中，我们使用了若干次矩方法，但矩何时可以决定分布函数？下面这个反例表明，并不是任何时候，矩都可以决定分布函数。

Example 2.1 (反例——对数正态分布). 先有一个密度函数：

$$f_0(x) = \frac{1}{\sqrt{2\pi}x} \exp\left(-\frac{1}{2}(\ln x)^2\right) \ (x > 0)$$

为标准对数正态分布的密度函数。再来一族分布函数：

$$f_a(x) = f_0(x)(1 + a\sin(2\pi \ln x)) \ (a \in (-1, 1))$$

可以发现这个密度函数算出来的各阶矩和 a 是没有关系的，所以在没有任何条件下，矩无法决定分布函数。

下面这个定理给出了在何种条件下矩可以决定分布函数。

Lemma 2.2 (矩收敛定理). 对于一个分布函数 $F(x)$，令 $\gamma_k = \int x^k \,\mathrm{d}F$，若：

$$\limsup_{k\to\infty} \frac{1}{2k} \gamma_{2k}^{\frac{1}{2k}} = r < \infty \tag{R}$$

则 $F(x)$ 被其矩决定，其中 (R) 被称为 Riesz 条件，是一个充分条件。

Proof. 令 $\mu_k = \int |x|^k \,\mathrm{d}F$，则有 Cauchy-Schwarz 不等式可得：

$$\mu_{2k+1} \le \sqrt{\mu_{2k}\mu_{2k+2}}$$

而 $\mu_{2k} = \gamma_{2k}$，这说明偶数阶矩给定后，绝对值的奇数阶矩也被控制，即：

$$\limsup_{k\to\infty} \frac{1}{k} \mu_k^{\frac{1}{k}} = r \tag{1}$$

利用带 Lagrange 余项的 Taylor 展开：

$$e^{it} - \sum_{m=0}^{n-1} \frac{(it)^m}{m!} = \frac{i^n}{(n-1)!} \int_0^t (t-s)^{n-1} e^{is} \,\mathrm{d}s$$

可得如下事实：

$$\left| e^{it} - \sum_{m=0}^{n-1} \frac{(it)^m}{m!} \right| \le \frac{|t|^n}{n!}, \ \forall t \in \mathbb{R}$$

于是：

$$\left|e^{i\theta X}\left(e^{itX}-\sum_{m=0}^{n-1}\frac{(itX)^m}{m!}\right)\right|\le\frac{|tX|^n}{n!}$$

对该式取期望就有：

$$\left|\phi(t+\theta)-\sum_{m=1}^{n-1}\frac{t^k}{m!}\phi^{(m)}(\theta)\right|\le\frac{t^n}{n!}\mu_n$$

再利用 (1) 与事实 $\frac{n^n}{n!}\le e^n$，可以得出：

$$\limsup_{n\to\infty}\mu_k^{\frac{1}{k}}=er$$

所以上面幂级数的收敛半径至少为 $\frac{1}{er}$，所以：

$$\phi(t+\theta)=\phi(\theta)+\sum_{m=1}^{\infty}\frac{(it)^k}{m!}\phi^{(m)}(\theta),\ \forall\theta\in\mathbb{R},|t|<\frac{1}{er}$$

取 $\theta=0$，则 $|t|<\frac{1}{er}$ 范围内的 $\phi(t)$ 可由 $\{\gamma_n\}$ 确定。再接着取 $\theta=\pm\frac{1}{e(r+1)},\ldots$，就可以在整个实轴上确定特征函数 $\phi(t)$，进而唯一确定分布函数 $F(x)$。 □

2.2 Lindeberg 替换术

Theorem 2.3 (Lindeberg-Feller 中心极限定理)**.** 对于一列相互独立的随机变量列（不一定同分布）$\{X_n\}$，有：$\mathbb{E}[X_k]=0,\ \sigma_k^2=\mathrm{Var}(X_k)<\infty,\ \forall k\in\mathbb{N}^+$，记 $B_n^2=\sum\limits_{i=1}^{n}\sigma_i^2$。

则称其满足 Lindeberg 条件，若：

$$\forall\epsilon>0,\ \lim_{n\to\infty}\frac{1}{B_n^2}\sum_{i=1}^{n}\mathbb{E}[X_i^2I_{\{|X_i|\ge\epsilon B_n\}}]=0 \tag{L}$$

若一列相互的随机变量列 $\{X_n\}$ 满足条件 (L)，则：

$$\frac{1}{B_n}\sum_{i=1}^{n}X_i\xrightarrow{D}\mathcal{N}(0,1) \tag{CLT}$$

且：

$$\lim_{n\to\infty}\frac{1}{B_n^2}\max_{1\le i\le n}\sigma_i^2=0 \tag{F}$$

其中 (F) 被称为 Feller 条件。

在本篇中，我们证明了 (L) $\Rightarrow$ (F)，但我们不知道怎么证明 (L) $\Rightarrow$ (CLT)，事实上有 (L) $\iff$ (CLT) + (F)。

对于独立同的情况，可以非常直接的使用若干相同特征函数的次幂来解决。对于相互独立但是不一定同分布的情况，同样我们也可以用特征函数来解决，但是处理必须非常小

心，可以参见 Rick Durrett *Probability: Theory and Example* 和钟开莱《概率论教程》上的证明。

这里我们介绍另一种证明 Lindeberg-Feller CLT 的方法，叫做 **Lindeberg 替换术**。

动机： 如果每个 X_n 都是服从某个正态分布的，则可以直接做（独立正态分布的可加性）。Lindeberg 替换术就是尝试将每个变量依次换成某个正态分布随机变量，并且控制误差，最后得到结果。

先证明引理，这个引理给出了依分布收敛的若干等价刻画，为我们之后证明做准备。

Lemma 2.4 (依分布收敛的等价刻画). 对于一个随机变量列 $\{X_n\}$，与一个随机变量 X，我们有如下几条等价：

(1) $X_n \xrightarrow{D} X$。
(2) 对于任何有界连续函数 $g \in C_b(\mathbb{R})$，有 $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$。
(3) 对于任何有界一致连续函数 g，有 $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$。
(4) 固定 $k \in \mathbb{N}$，对于任何 $0 \sim k$ 阶导都有界连续的函数 g（即 $g^{(i)} \in C_b(\mathbb{R})$, $\forall i = 0, 1, \ldots, k$），有 $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$。
(5) 特征函数逐点收敛，即：$\forall t \in \mathbb{R}$, $\phi_n(t) \to \phi(t)$。

Proof. 1、5 等价性可以使用 Lévy-Cramér 连续性定理来证明，而 1、2 的等价性可以使用 Skorokhod 表示定理来证明。又可以发现 3、4 是 2 的特殊情况，所以可以得到 $1 \Rightarrow 3, 4$。

又注意到：$\phi_n(t) = \mathbb{E}[\cos(tX_n)] + i\,\mathbb{E}[\sin(tX_n)]$, $\phi(t) = \mathbb{E}[\cos(tX)] + i\,\mathbb{E}[\sin(tX)]$。

而且 $\sin(tx), \cos(tx)$ 这两族函数都满足 3、4 中对函数的要求，所以如果以 3、4 为前提，就可以得出 $\forall t \in \mathbb{R}$, $\phi_n(t) \to \phi(t)$。

这样我们就可以证明 $1 \Rightarrow 3, 4 \Rightarrow 5 \Rightarrow 1$，进而这五个都等价。 □

接下来我们证明 Lindeberg-Feller 中心极限定理：

Proof of Theorem 2.3. 取 $\{Y_k\}$ 相互独立，且 $Y_k \sim \mathcal{N}(0, \sigma_k^2)$，而且还有 $\{Y_k\}, \{X_k\}$ 相互独立（这可以做到，由高等概率论中的 Kolmogorov 扩展定理）。

固定 n，对于任何 $0 \sim 3$ 阶导都有界连续的函数 g，令：

$$\zeta_{nk} = \sum_{i=1}^{k-1} X_i + \sum_{i=k+1}^{n} Y_i$$

则有 $\zeta_{nn} + X_n = \sum_{i=1}^{n} X_i, \zeta_{n1} + Y_1 = \sum_{i=1}^{n} Y_i$，而且 $\zeta_{nk} + X_k = \zeta_{n,k+1} + Y_{k+1}$, $\forall k \in [n-1]$。

因此，令 $Y \sim \mathcal{N}(0,1)$，则 Y 与 $\dfrac{1}{B_n} \sum_{i=1}^{n} Y_i$ 同分布，故错位相减可得：

$$\mathbb{E}\left[g\left(\frac{S_n}{B_n}\right)\right] - \mathbb{E}\left[g\left(Y\right)\right] = \sum_{k=1}^{n} \left(\mathbb{E}\left[g\left(\frac{\zeta_{nk} + X_k}{B_n}\right)\right] - \mathbb{E}\left[g\left(\frac{\zeta_{nk} + Y_k}{B_n}\right)\right]\right)$$

对于 $g(x)$，在 $\dfrac{\zeta_{nk}}{B_n}$ 这一点做 Taylor 展开可以得到：

$$g\left(\frac{\zeta_{nk} + X_k}{B_n}\right) = g\left(\frac{\zeta_{nk}}{B_n}\right) + g'\left(\frac{\zeta_{nk}}{B_n}\right) \cdot \frac{X_k}{B_n} + \frac{1}{2} g''(\cdot) \cdot \frac{X_k^2}{B_n^2} + \frac{1}{6} g'''(\cdot) \cdot \frac{X_k^3}{B_n^3} + \cdots$$

对于 Y_k 也有类似表达式。

因为对于任意 k，都有 ζ_{nk} 与 X_k,Y_k 独立，故先固定 ζ_{nk}，关于 X_k,Y_k 求期望，再配合 X_k,Y_k 一阶矩二阶矩都相同的条件可以得到：

$$\mathbb{E}\left[g'\left(\frac{\zeta_{nk}}{B_n}\right)\cdot\frac{X_k-Y_k}{B_n}\right]=\mathbb{E}\left[g''\left(\frac{\zeta_{nk}}{B_n}\right)\cdot\frac{X_k^2-Y_k^2}{B_n^2}\right]=0$$

有了这个我们可以发现 Taylor 展开式中一阶二阶项都是 0。

而令 $h(t)=\sup\limits_{x\in\mathbb{R}}\left|g(x+t)-g(x)-g'(x)t-\dfrac{1}{2}g''(x)t^2\right|$，因为 $g,g',g'',g'''\in C_b(\mathbb{R})$，所以 sup 里面的东西可以被放缩成一个系数全是有限常数的二次多项式，故存在常数 $K>0$，使得 $h(t)\le K\min\{t^2,|t|^3\}$。

进而：

$$\left|\mathbb{E}\left[g\left(\frac{S_n}{B_n}\right)\right]-\mathbb{E}\left[g\left(Y\right)\right]\right|\le\sum_{k=1}^{n}\left(\mathbb{E}\left[h\left(\frac{X_k}{B_n}\right)\right]+\mathbb{E}\left[h\left(\frac{Y_k}{B_n}\right)\right]\right)$$

先看 Y_k 部分，有：

$$\begin{aligned}\sum_{k=1}^{n}\mathbb{E}\left[h\left(\frac{Y_k}{B_n}\right)\right]&\le K\sum_{i=1}^{n}\mathbb{E}\left[\frac{|Y_k|^3}{B_n^3}\right]\ (h(t)\le K|t|^3)\\&=K\sum_{k=1}^{n}\frac{\sigma_k^3}{B_n^3}\mathbb{E}[|Y|^3]\ (\mathbb{E}[|Y_k|^3]=\sigma_k^3\,\mathbb{E}[|Y|^3])\\&\le\frac{K}{B_n}\max_{1\le k\le n}\sigma_k\cdot\sum_{k=1}^{n}\frac{\sigma_k^2}{B_n^2}\mathbb{E}[|Y|^3]\ (\text{选一个最大的}\sigma_k^2)\\&=K\sqrt{\frac{1}{B_n^2}\max_{1\le k\le n}\sigma_k^2}\cdot\mathbb{E}[|Y|^3]\ \left(B_n^2=\sum_{k=1}^{n}\sigma_k^2\right)\end{aligned}$$

可以看到最后一行根号里面的就是 Feller 条件的表达式，而 $Y\sim\mathcal{N}(0,1)$，故 $\mathbb{E}[|Y|^3],K$ 都是正常数，所以令 $n\to\infty$ 取极限就有：

$$\sum_{k=1}^{n}\mathbb{E}\left[h\left(\frac{Y_k}{B_n}\right)\right]\to 0\ (n\to\infty)$$

再来看 X_k 部分，有：

$$\begin{aligned}\sum_{k=1}^{n}\mathbb{E}\left[h\left(\frac{X_k}{B_n}\right)\right]&=\sum_{k=1}^{n}\mathbb{E}\left[h\left(\frac{X_k}{B_n}\right)I_{\{|X_k|<\epsilon B_n\}}\right]+\sum_{k=1}^{n}\mathbb{E}\left[h\left(\frac{X_k}{B_n}\right)I_{\{|X_k|\ge\epsilon B_n\}}\right]\\&\le\sum_{k=1}^{n}K\,\mathbb{E}\left[\frac{|X_k|^3}{B_n^3}I_{\{|X_k|<\epsilon B_n\}}\right]+\sum_{k=1}^{n}K\,\mathbb{E}\left[\frac{X_k^2}{B_n^2}I_{\{|X_k|\ge\epsilon B_n\}}\right]\ (h(t)\le K\min\{t^2,|t|^3\})\\&\le\sum_{k=1}^{n}\frac{K\epsilon}{B_n^2}\mathbb{E}[X_k^2]+\sum_{k=1}^{n}\frac{K}{B_n^2}\mathbb{E}\left[X_k^2I_{\{|X_k|\ge\epsilon B_n\}}\right]\ (\text{提出一个}X_k\text{放缩成}\epsilon B_n)\\&=K\epsilon+\frac{K}{B_n^2}\sum_{k=1}^{n}\mathbb{E}\left[X_k^2I_{\{|X_k|\ge\epsilon B_n\}}\right]\ \left(B_n^2=\sum_{k=1}^{n}\mathbb{E}[X_k^2]\right)\end{aligned}$$

先令 $n \to \infty$ 取上极限，发现最后一行右边项就是 Lindeberg 条件表达式，趋于 0。再令 $\epsilon \to 0$ 得到左边项也趋于 0，故：

$$\sum_{k=1}^{n} \mathbb{E}\left[h\left(\frac{X_k}{B_n}\right)\right] \to 0\ (n \to \infty)$$

将 X_k, Y_k 两个和式拼在一起即可得到：

$$\mathbb{E}\left[g\left(\frac{S_n}{B_n}\right)\right] \to \mathbb{E}\left[g\left(Y\right)\right]\ \ (n \to \infty)$$

故由 **Lemma** 2.4，可得 $\dfrac{S_n}{B_n} \xrightarrow{D} Y \sim \mathcal{N}(0,1)$。 □

3 统计物理

3.1 信息熵

Example 3.1 (投骰子)**.** 在赌场中连续投三个骰子，我们看如下两个事件：

- 出现三个六，概率为 $\dfrac{1}{6^3}$。
- 三个骰子点数之和为偶数，概率为 $\dfrac{1}{2}$。

相较于第二个事件，第一个事件发生概率小很多。如果第一个事件发生，人们脸上的表情会比发生第二个事件时惊讶很多。

我们想找到一个有关于概率 p 的函数 $S(p)$，可以刻画一个事件的惊奇程度。

Definition 3.2 (信息熵)**.** 称一个函数 $S: [0,1] \to \mathbb{R}$ 是一个信息熵函数，若其满足如下性质：

(1) $S(1) = 0$，这表示一个必然事件并不惊奇。

(2) 当 $p_1 < p_2$ 时，$S(p_1) > S(p_2)$，表示约不可能发生的事件，其惊奇程度越高，蕴含的信息越多。

(3) $S(p)$ 为关于 p 的连续函数。

(4) $S(pq) = S(p) + S(q)$。这表明若两个事件 A, B 独立，且 $\Pr(A) = p, \Pr(B) = q$，则 $A \cap B$ 的惊奇程度为 $S(pq)$。若先有 A 发生，再有 B 发生，那增量 $S(pq) - S(p)$ 应该与 A 无关，对 B 也同理。

实际上满足这几条的函数是唯一的。

Theorem 3.3 (信息熵函数的唯一性)**.** 信息熵函数 $S(p)$ 必然为：

$$S(p) = -c \ln p$$

其中 $c > 0$ 为一常数。

Proof. 由 4 知：

$$S(p^{m/n}) = \frac{m}{n}S(p)$$

再由 3 可知：

$$S(p^x) = xS(p),\ \forall x > 0$$

令 $x = -\ln p$ 就有：

$$S(p) = -S(1/e)\ln p$$

令 $c = S(1/e)$ 即得证。 □

我们再给出关于随机变量定义的熵——Shannon 熵。

Definition 3.4 (Shannon 熵)**.** 对于一个离散型随机变量 X。我们定义其 Shannon 熵为：

$$\mathrm{H}(X) = -\sum_i p_i \ln p_i$$

对于两个离散随机变量 X, Y，我们还有联合熵：

$$\mathrm{H}(X,Y) = -\sum_{i,j} p(x_i, y_j)\ln p(x_i, y_j)$$

与 X 关于 Y 的相对熵：

$$\mathrm{H}_Y(X) = \sum_j H_{Y=y_j}(X) p_Y(y_j)$$

其中：

$$\mathrm{H}_{Y=y_j}(X) = -\sum_i p(x_i \mid y_j)\ln p(x_i \mid y_j)$$

这里省略了对分布列的说明。

参考文献：1948, Shannon, ”A Mathematical Theory of Communications”.

对于 Shannon 熵，我们有如下两个重要事实：联合熵等于相对熵与原熵的加和，以及相对熵不超过原熵。证明较为简单，我们略去。

Lemma 3.5. 我们有如下两个事实：

- $\mathrm{H}(X,Y) = \mathrm{H}_Y(X) + \mathrm{H}(Y)$。
- $\mathrm{H}_Y(X) \le \mathrm{H}(X)$，且等号成立当且仅当 X 与 Y 独立。

Remark. 当随机变量 X 可取有限个值 $\Pr(X = x_i) = p_i,\ \forall i \in [n]$，则易知当且仅当 $p_i = \frac{1}{n},\ \forall i \in [n]$ 时，$\mathrm{H}(X)$ 取到最大值 $\ln n$。

而 $\ln n > \ln(n-1)$，依此我们可以发现：**不确定性程度越大，熵越大，熵是刻画混乱程度的量**。

Definition 3.6 (连续型随机变量的 Shannon 熵)**.** 设连续型随机变量 X 有密度函数 $f(x)$，则有 Shannon 熵：

$$\mathrm{H}(X) = -\int_{\mathbb{R}} f(x) \ln f(x) \, \mathrm{d}x$$

而对于两个连续型随机变量 X, Y，也有联合熵：

$$\mathrm{H}(X, Y) = -\iint_{\mathbb{R}^2} f(x, y) \ln f(x, y) \, \mathrm{d}x \, \mathrm{d}y$$

为了找出连续型随机变量熵的性质，我们先给出如下不等式。利用事实 $\ln \leq x - 1$，可以非常容易验证该不等式，这里略去不证。

Lemma 3.7 (Gibbs 不等式)**.** 我们有：

$$u - u \ln u \leq v - u \ln v, \ \forall u, v \in \mathbb{R}^+$$

我们取 $u = f(x), v = g(x)$，再积分就有：

$$\mathrm{H}(x) = -\int_{\mathbb{R}} f(x) \ln f(x) \, \mathrm{d}x \leq -\int_{\mathbb{R}} f(x) \ln g(x) \, \mathrm{d}x$$

借助信息熵，我们可以从另一个侧面，看正态分布在连续型分布中的特殊地位。

Theorem 3.8 (正态分布熵最大)**.** 记 $D_f = \{x \in \mathbb{R} \mid f(x) > 0\}$。对于一个连续型随机变量 X，其有密度函数 $f(x)$，则：

- 若 $D_f = \mathbb{R}$，且 $\mathbb{E}[X] = 0, \ \mathrm{Var}(X) = 1$。那么 $\mathrm{H}(X) \leq \ln \sqrt{2\pi e}$，且等号成立当且仅当 $X \sim \mathcal{N}(0, 1)$。
- 若 $D_f = (0, +\infty)$，且 $\mathbb{E}[X] = \dfrac{1}{\lambda} \ (\lambda > 0)$，则 $\mathrm{H}(X) \leq \ln\left(\dfrac{e}{\lambda}\right)$，且等号成立当且仅当 $X \sim \mathrm{Exp}(\lambda)$。
- 若 $D_f = (0, a)$，则 $\mathrm{H}(X) \leq \ln a$，且等号成立当且仅当 $X \sim U(0, a)$。

Proof. 我们只证明第一条，剩下几条证明类似。

我们取 $g(x) = \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{1}{2}x^2\right)$，则由 Gibbs 不等式：

$$\mathrm{H}(X) \leq -\int_{\mathbb{R}} f(x) \left(-\frac{1}{2}x^2 - \ln \sqrt{2\pi}\right) = \frac{1}{2} + \ln \sqrt{2\pi}$$

而当且仅当并不好说明，我们略去。 □

我们还有离散情形的 Gibbs 不等式：

Lemma 3.9. 若 $\forall i \in [n], \ p_i, q_i \geq 0$，且：

$$\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i,$$

则：

$$-\sum_{i=1}^{n} p_i \ln p_i \leq -\sum_{i=1}^{n} p_i \ln q_i$$

且等号成立当且仅当 $\forall i \in [n], \ p_i = q_i$。

由离散情形的 Gibbs 不等式，我们可以引出另一种熵——Boltzmann 熵：

Problem 3.10 (Boltzmann 熵). 对于一个离散型随机变量 X，是一个代表能级的物理量。其有分布列：$\Pr(X=E_i)=p_i,\ \forall i\in[n]$。

给定平均能量 $\mathbb{E}[X]=\sum\limits_{i=1}^{n}p_iE_i=U$，问 $\{p_i\}$ 取何值时其熵 $\mathrm{H}(X)$ 最大。

Solution. 取 $q_i=\dfrac{1}{Z}e^{-\beta E_i}$，其中 $Z=\sum\limits_{i=1}^{n}e^{-\beta E_i}$，$\beta>0$。

由离散情形的 Gibbs 不等式：

$$\mathrm{H}(X)\le-\sum_{i=1}^{n}p_i(-\beta E_i-\ln Z)=\beta U+\ln Z$$

在物理上 β 可由 $\{E_i\}$ 和 U 决定。注意对比只给定 n，不给定 U 的情况。 △

由此可以给出物理上的 Boltzmann 熵：

Definition 3.11 (Boltzmann 熵). $S=k_B\ln\Omega$,其中 Ω 为微观粒子状态数,k_B 为 Boltzmann 常数。

3.2 Ising 模型

Example 3.12 (格点气体). 设 $\Lambda\subset\mathbb{Z}^d$ 且 $N=|\Lambda|<\infty$，是 d 维格点的有限子集，而样本空间为 $\Omega_\Lambda=\{+1,-1\}^\Lambda$。

引入 Hamilton 量：

$$H(\omega)=-\sum_{i,j\in\Lambda}J_{ij}\omega_i\omega_j,\ \forall\omega\in\Omega_\Lambda$$

其中 $\{J_{ij}\}$ 给定，表示粒子各个分量相互作用的强度。由此我们可以定义 Ω_Λ 上的概率测度：

$$\mu_{\Lambda\beta}=\frac{1}{Z_{\Lambda\beta}}e^{-\beta H(\omega)}$$

其中 $Z_{\Lambda\beta}$ 是一个规范化常数。这个是统计物理里的格点气体模型。

更一般的，我们可以根据统计力学的基本假设，以及最大熵原理，引入 Gibbs 测度：

Definition 3.13 ((正则) Gibbs 分布). 我们有统计力学的基本假设：

(1) 微观状态空间 Ω_Λ，记 $N=|\Lambda|<\infty$ 为微观粒子状态数；

(2) 相互作用的 Hamilton 量 $H:\Omega_\Lambda\to\mathbb{R}$。

则我们可以定义 Gibbs 测度：

$$\mu_{\Lambda\beta N}(\omega)=\frac{1}{Z_{\Lambda\beta N}}e^{-\beta H(\omega)},\ \forall\omega\in\Omega_\Lambda$$

这里 $\beta\ge0$ 为逆温度 $\dfrac{1}{k_BT}$，ω 的 Boltzmann 权重为 $e^{-\beta H(\omega)}$，而配分函数：

$$Z_{\Lambda\beta N}=\sum_{\omega\in\Omega_\Lambda}e^{-\beta H(\omega)}$$

对于这一统计力学模型，我们关心如下问题：

(1) 热力学极限，$N \to \infty$。

(2) 相变现象。

(3) 简单逼近复杂。

接下来我们引入 Ising 模型。

在 1895 年，Pierre Curie 发现，加热一个磁铁，到达一定临界温度，磁铁的磁性会消失。这一临界温度被称为 Curie 温度。

1920 年 Heinrich Lenz 和 1925 年 Ernst Ising 提出如下模型来解释这一现象。

Definition 3.14 (Ising 模型). $\Lambda \subset \mathbb{Z}^d$ 为有限格点集合，构形空间 $\Omega_\Lambda = \{+1, -1\}^d$。

定义相邻点对集合：

$$\varepsilon_\Lambda = \{(i,j) \subset \Lambda^2 \mid \|i-j\|_1 = 1\}$$

$$\varepsilon_{\Lambda_L} = \{(i,j) \subset \Lambda_L^2 \mid \|i-j\|_1 \equiv 1 \pmod L\}$$

这里 $\Lambda_L = [L]^d$，在第二个集合里 $1, L$ 也是相邻的。

对于一个粒子 $\sigma \in \Omega_\Lambda$，记 $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)$。那么有 Ising 模型：

(1) **自由边界**：

$$H(\sigma) = -J \sum_{(i,j)\in\varepsilon_\Lambda} \sigma_i\sigma_j - h\sum_{i\in\Lambda}\sigma_i$$

(2) **周期边界**：

$$H(\sigma) = -J \sum_{(i,j)\in\varepsilon_{\Lambda_L}} \sigma_i\sigma_j - h\sum_{i\in\Lambda}\sigma_i$$

J 表示相互作用强度，而 h 表示外场强度。两种情况都可以定义对应的 Gibbs 测度。

我们关心**磁化强度** $M_\Lambda(\sigma) = \sum_{i\in\Lambda}\sigma_i$，在平均下的极限行为 $\dfrac{M_\Lambda(\sigma)}{N} \in [-1,1],\ N \to \infty$；以及在温度 T 变化（β 相应变化）时 $M_\Lambda(\sigma)$ 的涨落行为（比如当 $T = \infty, \beta = 0$ 时，Boltzmann 权重都是 1，就是一个古典概型；而当 $T \to 0, \beta \to +\infty$，$H(\sigma)$ 非常小，能量与熵相互竞争，相变可能出现）。

以 $d = 1$ 周期边界这一情况来阐述 Ising 模型（$d \geq 3$ 时非常困难），我们可以写出 Hamilton 量和配分函数：

$$H(\sigma) = -J\sum_{i=1}^N \sigma_i\sigma_{i+1} - h\sum_{i=1}^N \sigma_i$$

$$Z_{\Lambda\beta h} = \sum_{\sigma_1,\ldots,\sigma_N=\pm1} \exp\left(\beta J\sum_{i=1}^N \sigma_i\sigma_{i+1} + \beta h\sum_{i=1}^N \sigma_i\right)$$

约定 $\sigma_{N+1} = \sigma_1$。

Definition 3.15 (转移矩阵). 我们引入转移矩阵 P：

$$\langle\sigma|P|\sigma'\rangle = \exp\left(\beta J\sigma\sigma' + \frac{1}{2}\beta h(\sigma+\sigma')\right)$$

其中 Dirac 记号 $\langle\sigma|P|\sigma'\rangle$ 表示矩阵 P 在 (σ,σ') 处元素。具体展开写有：

$$P = \begin{array}{c c} & \begin{array}{cc} + & - \end{array} \\ \begin{array}{c} + \\ - \end{array} & \begin{pmatrix} e^{\beta J+\beta h} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta J-\beta h} \end{pmatrix} \end{array}$$

这样配分函数可以写为：

$$Z_{\Lambda\beta h} = \mathrm{tr}\big(P^N\big) = e^{N\beta J}(A_+^N(h) + A_-^N(h))$$

其中：

$$A_\pm(h) = \cosh(\beta h) \pm \sqrt{\sinh^2(\beta h) + e^{-4\beta J}}$$

并且 $\lambda_\pm = e^{\beta J}A_\pm(h)$ 为 P 的两个特征值。

进而：

$$\begin{aligned}\mathbb{E}\left[\frac{M_N(\sigma)}{N}\right] &= \frac{1}{\beta N}\frac{\partial}{\partial h}\ln Z_{\Lambda\beta h} = \frac{1}{\beta N}\frac{\partial}{\partial h}\ln\big(A_+^N(h) + A_-^N(h)\big) \\ &\to \frac{1}{\beta}\frac{\partial}{\partial h}\ln A_+(h) = \frac{\sinh(\beta h)}{\sqrt{\sinh^2(\beta h) + e^{-4\beta J}}}\end{aligned}$$

这个是在期望意义下磁化强度的极限行为。实际上我们对于 Ising 模型也有 LLN 和 CLT。

Theorem 3.16 (Ising 模型的 LLN 与 CLT)**.** 取 Ising 模型中 $d=1$, $h=0$，且周期边界，则：

$$\frac{M_N(\sigma)}{N} \xrightarrow{P} 0,\ \frac{M_N(\sigma)}{B\sqrt{N}} \xrightarrow{D} \mathcal{N}(0,1)\ (N\to\infty)$$

这里 $B = \sqrt{\dfrac{1+e^{2\beta J}}{1+e^{-2\beta J}}}$。

Proof. 我们考虑 $\dfrac{M_N(\sigma)}{N^\delta}$ 的矩母函数：

$$M(t) = \mathbb{E}\left[\exp\left(t\frac{M_N(\sigma)}{N^\delta}\right)\right] = \frac{Z_{\Lambda\beta h}}{Z_{\Lambda\beta 0}},\ h = \frac{t}{\beta N^\delta},\ \forall t\in\mathbb{R}$$

而：

$$\ln\left(\frac{Z_{\Lambda\beta h}}{Z_{\Lambda\beta 0}}\right) = \ln\left(\frac{A_+^N(h) + A_-^N(h)}{A_+^N(0) + A_-^N(0)}\right)$$

我们将 $A_\pm(h)$ 在 0 处 Taylor 展开：

$$A_\pm(h) = 1 \pm e^{-2\beta J} + \frac{1}{2}(1 \pm e^{2\beta J})(\beta h)^2 + O((\beta h)^4)$$

所以；

$$\begin{aligned}\ln\left(\frac{Z_{\Lambda\beta h}}{Z_{\Lambda\beta 0}}\right) &= \ln\left(\frac{A_+^N(h) + A_-^N(h)}{A_+^N(0) + A_-^N(0)}\right) \\ &= N\ln\left(1 + \frac{1}{2}\frac{1+e^{2\beta J}}{1+e^{-2\beta J}}(\beta h)^2 + O((\beta h)^4)\right)\end{aligned}$$

$$= N(\beta h)^2 \frac{1+e^{2\beta J}}{2(1+e^{-2\beta J})} + O(N(\beta h)^4)$$

当 $\delta = 1$ 时，极限为 0；当 $\delta = \frac{1}{2}$ 时，极限为 $\frac{B^2}{2}t^2$。这两个取一个 exp 分别对应零分布和 $\mathcal{N}(0, B^2)$ 的矩母函数，故得证。 □

这样在 $d = 1$ 周期边界条件下的事情已经明了了。当 $d = 2$ 时，LLN 显得稍微有点不同：

$$\mathbb{E}\left(\left|\frac{M_N(\sigma)}{N}\right|\right) \to \begin{cases} 0 & T > T_c \\ (1-\sinh^{-4}(2\beta J))^{1/8} & T \leq T_c \end{cases}$$

可以看到在温度小于 Cuire 温度 T_c 时，收敛到的值不一样，这个是相变现象一种概率诠释。

3.3 Cuire-Weiss 模型和相变现象

接下来我们介绍 Cuire-Weiss 模型，一定程度上可以说是 Ising 模型的平均场近似。在这一模型下，我们可以研究相变现象。

Definition 3.17 (Cuire-Weiss 模型). 一定程度上可以说是 Ising 模型的平均场近似，我们有：

$$\sum_{j:j\sim i} \sigma_i\sigma_j = 2d\sigma_i \frac{1}{2d}\sum_{j:j\sim i}\sigma_j \to \frac{2d\sigma_i}{N}\sum_{j=1}^N \sigma_j$$

我们使用整体平均，近似 Hamilton 量为：

$$H(\sigma) = -\frac{dJ}{N}\left(\sum_{i=1}^N \sigma_i\right)^2 - h\sum_{i=1}^N \sigma_i$$

这时候也定义相应的 Gibbs 测度为：

$$\mu_{N\beta h}(\sigma) = \frac{1}{Z_{N\beta h}} e^{-\beta H(\sigma)}$$

Theorem 3.18 (Cuire-Weiss 模型的相变现象). 令 $J = 1, h = 0, \beta_c = \frac{1}{2d}$，则：

(1) 当 $0 < \beta \leq \beta_c$ 时，对于任何的 $\epsilon > 0$，存在依赖于 β, ϵ 的常数 $c = c(\beta, \epsilon) > 0$，使得对充分大的 N，有：

$$\mu_{N\beta h}\left(\frac{M_N(\sigma)}{N} \notin (-\epsilon, \epsilon)\right) \leq 2e^{-cN}$$

(2) 当 $\beta > \beta_c$ 时，存在某一依赖于 β 的常数（自发磁化强度）$m^* = m^*(\beta) > 0$，使得对于任意充分小的 $\epsilon > 0$，存在一个依赖于 β, ϵ 的常数 $b = b(\beta, \epsilon) > 0$，当 N 充分大时有：

$$\mu_{N\beta 0}\left(\frac{M_N(\sigma)}{N} - m^* \notin (-\epsilon, \epsilon) \text{ and } \frac{M_N(\sigma)}{N} + m^* \notin (-\epsilon, \epsilon)\right) \leq 2e^{-bN}$$

为了证明这一定理，我们需要如下引理：

Lemma 3.19. 令自由能 $f_\beta(m) = -\beta dm^2 - S(m)$ ($m \in [-1,1]$)，这里 $S(m)$ 为熵：

$$S(m) = -\frac{1-m}{2}\ln\left(\frac{1-m}{2}\right) - \frac{1+m}{2}\ln\left(\frac{1+m}{2}\right)$$

则：

$$\lim_{N\to\infty}\frac{1}{N}\ln Z_{N\beta 0} = -\min_{m\in[-1,1]} f_\beta(m)$$

更一般的，对 $K \subset [-1,1]$，有：

$$\lim_{N\to\infty}\frac{1}{N}\ln \mu_{N\beta 0}\left(\frac{M_N(\sigma)}{N}\in K\right) = -\min_{m\in K} I_\beta(m)$$

这里 $I_\beta(m) = f_\beta(m) - \min_{m'\in[-1,1]} f_\beta(m')$，为速率函数，在大偏差理论中很有用。

Proof. 记 $A_N = \left\{-1+\frac{2k}{N}\;\middle|\; k=0,1,\ldots,N\right\}$，则由二项式的组合意义知：

$$\mu_{N\beta h}\left(\frac{M_N(\sigma)}{N}\in K\right) = \sum_{m\in K\cap A_N}\mu_{N\beta 0}\left(\frac{M_N(\sigma)}{N} = m\right)$$

$$= \frac{1}{Z_{N\beta h}}\sum_{m\in K\cap A_N}\binom{N}{(1+m)N/2}e^{\beta dm^2N+\beta hmN}$$

类似的，有：

$$Z_{N\beta h} = \sum_{m\in A_N}\binom{N}{(1+m)N/2}e^{\beta dm^2N+\beta hmN} \tag{2}$$

我们只证 $h=0$ 情况的 (2)，剩下部分证明类似。做一个粗放缩：

$$\Delta = \max_{m\in A_N}\binom{N}{(1+m)N/2}e^{\beta dm^2N},\ \Delta \le Z_{N\beta 0} \le (N+1)\Delta$$

再来，利用事实：存在常数 $c_1, c_2 > 0$，使得：

$$\frac{c_1}{\sqrt{N}}e^{NS(m)} \le \binom{N}{(1+m)N/2} \le c_2\sqrt{N}e^{NS(m)}$$

我们可以对 $Z_{N\beta 0}$ 做上下界估计。

先做上界估计：

$$\begin{aligned} Z_{N\beta 0} &\le c_2\sqrt{N}(N+1)\exp\left(N\max_{m\in A_N}\{\beta dm^2 + S(m)\}\right) \\ &\le c_2\sqrt{N}(N+1)\exp\left(-N\min_{m\in[-1,1]} f_\beta(m)\right) \end{aligned}$$

故：

$$\limsup_{N\to\infty}\frac{1}{N}\ln Z_{N\beta 0} \le -\min_{m\in[-1,1]} f_\beta(m)$$

再做下界估计：由 $f_\beta(m)$ 连续性知，存在 $m' \in [-1,1]$，使得 $f_\beta(m') = \min_{m\in[-1,1]} f_\beta(m)$。

则对于任意 $\delta > 0$，可取充分大的 N，使得存在 $m \in A_N$，满足：

$$|f_\beta(m) - f_\beta(m')| \le \delta$$

因此：

$$Z_{N\beta 0} \ge \frac{c_1}{\sqrt{N}} \exp(-N(f_\beta(m') + \delta))$$

进而：

$$\liminf_{N\to\infty} \frac{1}{N} \ln Z_{N\beta 0} \ge -f_\beta(m') - \delta$$

由 δ 的任意性，令 $\delta \to 0+$ 即得证（实际上这时候 N 也跟着一起趋向于 ∞）。 □

接下来我们证明原定理：

***Proor of Theorem** 3.18.* 我们讨论 $I_\beta(m) = f_\beta(m) - \min_{m'\in[-1,1]} f_\beta(m')$ 在 $[-1,1]$ 的最小值点（最小值为 0），实际上看的是 $f_\beta(m)$ 的拐点。

对 $f(m) = f_\beta(m)$ 求一阶和二阶导：

$$f'(m) = -2\beta dm + \frac{1}{2} \ln\left(\frac{1+m}{1-m}\right)$$

$$f''(m) = -2\beta d + \frac{1}{1-m^2}$$

我们分两种情况讨论：

(1) 当 $2\beta d \le 1$ 时，则 $f''(m) \ge 0$ 恒成立，而 $f'(m) = 0$ 当且仅当 $m = 0$，这也是 $I_\beta(m)$ 的唯一最小值点。

我们取 $K = [-1,-\epsilon] \cup [\epsilon, 1]$，则在 K 上，$I_\beta(m)$ 的最小值是一个和 ϵ, β 有关的常数，记为 $c = c(\beta,\epsilon) > 0$。故由 **Lemma** 3.19：

$$\lim_{N\to\infty} \frac{1}{N} \ln \mu_{N\beta 0}\left(\frac{M_N(\sigma)}{N} \in K \ (\notin (-\epsilon,\epsilon))\right) = -c$$

故：

$$\mu_{N\beta 0}\left(\frac{M_N(\sigma)}{N} \notin (-\epsilon,\epsilon)\right) \le 2e^{-cN}$$

其中右边乘 2 是为了包括取极限带来的误差。

(2) 当 $2\beta d > 1$ 时，$f''(m) = 0$ 有两根 $m = \pm\sqrt{1 - \frac{1}{2\beta d}} \triangleq \pm m^*$，这也是 $I_\beta(m)$ 在 $[-1,1]$ 上的两个最小值点。

我们取 $K = [-1,1] \backslash \big((m^* - \epsilon, m^* + \epsilon) \cup (-m^* - \epsilon, -m^* + \epsilon)\big)$，之后和上一个情况完全类似。

综上得证。 □

Remark. 通过 **Theorem** 3.18，我们可以得到一个有着相变现象的“LLN”：

$$\frac{M_N(\sigma)}{N} \xrightarrow{P} \begin{cases} 0 & \beta \le \beta_c \\ \frac{1}{2}(\delta(m^*) + \delta(-m^*)) & \beta > \beta_c \end{cases}$$

其中 $\delta(x)$ 为 Dirac 函数，$\dfrac{1}{2}(\delta(m^*)+\delta(-m^*))$ 表示以概率 $\dfrac{1}{2}$ 取到 $\pm m^*$ 的随机变量的密度函数。

有了“LLN”，我们自然还关心“CLT”，即 $\dfrac{M_N(\sigma)}{N^\delta}$ 的极限行为。

我们还是看矩母函数：

$$M(t)=\mathbb{E}\left[\exp\left(t\frac{M_N(\sigma)}{N^\delta}\right)\right]=\frac{Z_{N\beta h}}{Z_{N\beta 0}},\ h=\frac{t}{\beta N^\delta}$$

对于 $Z_{N\beta h}$，有：

$$Z_{N\beta h}=\sum_{\{\sigma\}}\exp\left(\frac{\beta d}{N}\left(\sum_{i=1}^N\sigma_i\right)^2+\beta h\sum_{i=1}^N\sigma_i\right)$$

我们不喜欢那个关于和式的二次方项。为此，我们使用技巧 **Hubbard-Stratonovich 变换**：

$$e^{\alpha x^2}=\frac{1}{\sqrt{\pi\alpha}}\int_{\mathbb{R}}\mathrm{d}y\exp\left(-\frac{y^2}{\alpha}+2xy\right),\ \alpha>0$$

可得：

$$\begin{aligned}
Z_{N\beta h}&=\sum_{\{\sigma\}}\exp\left(\frac{\beta d}{N}\left(\sum_{i=1}^N\sigma_i\right)^2+\beta h\sum_{i=1}^N\sigma_i\right)\\
&=\sqrt{\frac{N}{\pi\beta d}}\int_{\mathbb{R}}\mathrm{d}y\sum_{\{\sigma\}}\exp\left(-\frac{N}{\beta d}y^2+(2y+\beta h)\sum_{i=1}^N\sigma_i\right)\\
&=\sqrt{\frac{N}{\pi\beta d}}\int_{\mathbb{R}}\mathrm{d}y\exp\left(-\frac{N}{\beta d}y^2\right)\left(e^{2y+\beta h}+e^{-2y-\beta h}\right)^N\\
&=\sqrt{\frac{N}{4\pi\beta d}}\int_{\mathbb{R}}\mathrm{d}y\exp\left(-\frac{\beta h^2N}{4d}+\frac{hN}{2d}y-Ng(y)\right)\ (\text{做代换}y\leftarrow(y-\beta h)/2)
\end{aligned}$$

其中 $g(y)=\dfrac{1}{4\beta d}y^2-\ln(e^y+e^{-y})$。代回去得到：

$$M(t)=\exp\left(-\frac{t^2}{4\beta dN^{2\delta-1}}\right)\cdot\frac{\displaystyle\int_{\mathbb{R}}\exp\left(\frac{ty}{2\beta dN^{\delta-1}}-Ng(y)\right)\mathrm{d}y}{\displaystyle\int_{\mathbb{R}}e^{-Ng(y)}\,\mathrm{d}y}$$

我们需要找到一种方法去处理分子分母里的积分，为此我们引入：

Example 3.20 (Laplace 方法)**.** 我们有欧拉积分：

$$\Gamma(s+1)=\int_0^{+\infty}x^se^{-x}\,\mathrm{d}x=s^{s+1}\int_0^{+\infty}e^{-sg(x)}\,\mathrm{d}x$$

其中最后一步做替换 $x\leftarrow sx$，而 $g(x)=x-\ln x$。

我们换一种方式来证明 Stirling 公式，大致上的想法就是，在 $s\to+\infty$ 时，上面的积分式里只有 $g(x)$ 的极小值点部分产生的贡献是主要的。

$g(x)$ 在 $x=1$ 处取到极小值 1，我们在 $x=1$ 这一点 Taylor 展开，配合 $g'(1)=0$ 有：

$$g(x)-g(1)=\frac{g''(1)}{2}(x-1)^2+O((x-1)^3)$$

O 里面的小项，在 exp 作用会变得"更小"，所以我们只需要看二阶导对应的主项：

$$\begin{aligned}&e^{-sg(1)}\int_0^{+\infty}e^{-s(g(x)-g(1))}\,\mathrm{d}x\\ \sim&e^{-s}\int_{1-\delta}^{1+\delta}\exp\left(-\frac{g''(1)}{2}(x-1)^2\right)\mathrm{d}x\\ \sim&\frac{e^{-s}}{\sqrt{s}}\int_{-\sqrt{s}\delta}^{\sqrt{s}\delta}\exp\left(-\frac{x^2}{2}\right)\mathrm{d}x\sim\sqrt{\frac{2\pi}{s}}e^{-s}\end{aligned}$$

其中最后一个 $\sim$ 是因为当 $s\to+\infty$ 时，那个积分会变成高斯积分。发现这时候再乘 s^{s+1} 就得到 Stirling 公式。

这种方法实际上叫 Laplace 方法，上面说的并不严谨，更多可以参见：Wikipedia: Laplace Method。

回到原问题，我们对 $g(y)$ 求一阶和二阶导：

$$g'(y)=\frac{y}{2\beta d}-\frac{e^y-e^{-y}}{e^y+e^y}$$

$$g''(y)=\frac{1}{2\beta d}-\frac{4e^{2y}}{(1+e^{2y})^2}$$

当 $\beta\le\beta_c$ 时，$g(y)$ 有唯一的最小值点 $y_0=0$。

还是根据 β 分类讨论：

(1) 当 $\beta<\beta_c$ 时，有 $g''(y_0)>0$。

将 $g(y)$ 在 y_0 这一点 Taylor 展开：

$$g(y)-g(y_0)=\frac{1}{2}\left(\frac{1}{2\beta d}-1\right)y^2+O(y^4)$$

还是仿照 Laplace 方法，只看主项，就有：

$$M(t)\sim\exp\left(-\frac{t^2}{4\beta d}\right)\frac{\int_{-\epsilon}^{\epsilon}\exp\left(\frac{t}{2\beta d}\sqrt{N}y-\frac{N}{2}\left(\frac{1}{2\beta d}-1\right)y^2\right)\mathrm{d}y}{\int_{-\epsilon}^{\epsilon}\exp\left(-\frac{N}{2}\left(\frac{1}{2\beta d}-1\right)y^2\right)\mathrm{d}y}$$

$$\sim\exp\left(-\frac{t^2}{4\beta d}+\frac{t^2}{4\beta d(2\beta d-1)}\right)\frac{\int_{-\sqrt{N}\epsilon}^{\sqrt{N}\epsilon}\exp\left(\left(\frac{t}{\sqrt{4\beta d(2\beta d-1)}}-\sqrt{\frac{1-\beta d}{4\beta d}}u\right)^2\right)\mathrm{d}u}{\int_{-\sqrt{N}\epsilon}^{\sqrt{N}\epsilon}\exp\left(-\frac{1}{2}\left(\frac{1}{2\beta d}-1\right)u^2\right)\mathrm{d}u}$$

$$\sim\exp\left(\frac{t^2}{2(1-2\beta d)}\right)\frac{\int_{\mathbb{R}}\exp\left(-\frac{N}{2}\left(\frac{1}{2\beta d}-1\right)u'^2\right)\mathrm{d}u'}{\int_{\mathbb{R}}\exp\left(-\frac{N}{2}\left(\frac{1}{2\beta d}-1\right)u^2\right)\mathrm{d}u}\sim\exp\left(\frac{t^2}{2(1-2\beta d)}\right)$$

其中第一行到第二行使用换元 $u=\sqrt{N}y$，并且将分子进行配方。第二行到第三行，先利用当 $N\to\infty$ 时，积分区域可以扩展到 $\mathbb{R}$，再进一步利用分子积分的平移不变性。

可以看到这时候 $M(t)$ 极限 $\mathcal{N}(0,B^2)$ 的矩母函数，其中 $B=\dfrac{1}{\sqrt{1-2\beta d}}$。

(2) 当 $\beta=\beta_c$ 时，$g''(y_0)=0$，这时候 Taylor 展开只看二次项不够了，要看四次项：

$$g(y)-g(y_0)=\frac{1}{12}y^4+O(y^6)$$

那么令 $\delta=\dfrac{3}{4}$，我们有：

$$M(t)\sim\exp\left(-\frac{t^2}{4\beta d\sqrt{N}}\right)\frac{\displaystyle\int_{-\epsilon}^{\epsilon}\exp\left(\frac{N^{1/4}t}{2\beta d}y-\frac{1}{12}y^4\right)\mathrm{d}y}{\displaystyle\int_{-\epsilon}^{\epsilon}\exp\left(-\frac{1}{12}y^4\right)\mathrm{d}y}\sim\frac{\displaystyle\int_{\mathbb{R}}\exp\left(tx-\frac{1}{12}x^4\right)\mathrm{d}x}{\displaystyle\int_{\mathbb{R}}\exp\left(-\frac{1}{12}x^4\right)\mathrm{d}x}$$

其中最后一步同前面类似：做了换元 $x\leftarrow N^{1/4}y$；然后再用：当 $N\to\infty$ 时，积分区域可以扩展到 $\mathbb{R}$；并且还用了：分式外的 $\exp\left(\cdot/\sqrt{N}\right)$ 项，当 $N\to\infty$ 时会趋向于 1。

可以看到这时候 $M(t)$ 极限是一个密度函数为 $\dfrac{1}{c}\exp\left(-\dfrac{1}{12}x^4\right)$ 的随机变量，其中 $c=\displaystyle\int_{\mathbb{R}}\exp\left(-\frac{1}{12}x^4\right)\mathrm{d}x$ 是一规范化常数。

综上所述，我们有结论：

Theorem 3.21 (Cuire-Weiss 模型的更多极限行为)**.** 令 $J=1,h=0,\beta_c=\dfrac{1}{2d}$，则：

(1) 当 $\beta<\beta_c$ 时：

$$\frac{M_N(\sigma)}{B\sqrt{N}}\xrightarrow{D}\mathcal{N}(0,1),\ B=\frac{1}{\sqrt{1-2\beta d}}$$

(2) 当 $\beta=\beta_c$ 时：

$$\frac{M_N(\sigma)}{N^{3/4}}\xrightarrow{D}\text{Some distribution with density }\frac{1}{c}\exp\left(-\frac{1}{12}x^4\right)$$

3.4 李政道-杨政宁单位圆定理

我们重写 Ising 模型的配分函数 $Z_{N\beta h}$，把 $\sigma_i\sigma_j,\sigma_i$ 中为 1 的那些抛去不看：

$$Z_{N\beta h}=\exp\left(\beta hN+\beta\sum_{i<j}J_{ij}\right)\sum_{\{\sigma\}}\exp\left(\beta h\sum_{i=1}^{N}(\sigma_i-1)+\beta\sum_{i<j}J_{ij}(\sigma_i\sigma_j-1)\right)$$

令 $z=e^{-2\beta h}$，$a_{ij}=e^{-2\beta J_{ij}}$，$X(\{\sigma\})=\{i\in[N]\mid\sigma_i=-1\}$，则有：

$$Z_{N\beta h}=\exp\left(\beta hN+\beta\sum_{i<j}J_{ij}\right)\sum_{X\subset[N]}z^{|X|}\prod_{i\in X,j\notin X}a_{ij}$$

我们进行一个推广，定义一个关于 N 个复变元 $z_1,z_2,\ldots,z_N$ 的多项式：

$$P(z_1,\ldots,z_N)=\sum_{X\subset[N]}z^X\prod_{i\in X,j\notin X}a_{ij}$$

其中 $z^X = \prod_{i \in X} z_i$。

关于这个多项式，我们可以刻画其零点的性质，这就是李政道-杨政宁单位圆定理。

Theorem 3.22 (李政道-杨政宁单位圆定理). 若对于任何 i, j 有 $a_{ij} \in [-1, 1]$，则当 $|z_i| < 1, \ldots, |z_N| < 1$ 时，$P(z_1, \ldots, z_N) \neq 0$。

特别的对于 $|z| < 1$，有 $P(z, \ldots, z) \neq 0$。进而，当 $|z| > 1$ 时，我们有 $P(1/z, \ldots, 1/z) = z^{-N} P(z, \ldots, z) \neq 0$。也就是说，$P(z, \ldots, z)$ 的零点只能在 $|z| = 1$ 这样一个单位圆上。

Proof. 考试不考，摆烂。 □

Remark. 在 $\mu_{N\beta 0}$ 下看 $Y = \beta \sum_{i=1}^{N} \sigma_i$ 其矩母函数：

$$M_Y(t) = \mathbb{E}[e^{tY}] = \frac{Z_{N\beta t}}{Z_{N\beta 0}} = 2^{-N/2} \frac{P_N(z)}{P_N(1)}$$

若 $h = it$，则 Y 的特征函数 $\phi_Y(t)$，只会有实零点（即使 t 取遍整个复数域 $\mathbb{C}$）。

再来，配分函数可能有实零点后，取对数就会有奇点。奇点和相变现象很有关系，没有奇点就没有相变，有奇点就有相变。

依此我们可以定义一类随机变量：

Definition 3.23 (李-杨类随机变量). 称一个随机变量 X 是李-杨类随机变量，若：

(1) X 是对成的，即 X 与 $-X$ 同分布。

(2) $\mathbb{E}[e^{bX^2}] < \infty$，对于某个 $b > 0$。

(3) 特征函数 $\phi(t)$ 只有实零点。

还有一个定理：

Theorem 3.24. 若一列随机变量 $\{X_n\}_{n=1}^{\infty}$ 都是李-杨类随机变量，且 $X_n \xrightarrow{D} X$，则 X 也是李-杨类随机变量。

参考文献： Charles M. Newman, Wei Wu. *Lee-Yang Property and Gaussian Multiplicative Chaos.* Communications In Mathematical Physics, 2019.

利用这一定理，我们先发现磁化强度 $M_N(\sigma) = \sum_{i=1}^{N} \sigma_i$ 是李-杨类随机变量。于是根据前面的定理，我们可以得到 $\mathcal{N}(0, 1)$ 和 $\frac{1}{c} \exp\left(-\frac{1}{12} x^4\right) \mathrm{d}x$ 都是李-杨类随机变量，后者在量子场论中很有用（ϕ^4 模型）。

李政道-杨政宁单位圆定理与 Riemann 猜想有一定联系。我们有 Riemann Zeta 函数：

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \ \mathrm{Re}(s) > 1$$

令 $\xi(z) = \frac{1}{2} s(s-1) \Gamma\left(\frac{s}{2}\right) \zeta(s) = \frac{1}{2} s(s-1) \Gamma\left(\frac{1-s}{2}\right) \zeta(1-s)$。则 $\xi(s) = \xi(-s)$，且为整函数。

利用 Fourier 变换，可得：

$$\xi(s) = \int_{\mathbb{R}} e^{isu}\Phi(u)\,\mathrm{d}u,\ \Phi(u) = \sum_{n=1}^{\infty}\left(4\pi^2 n^4 e^{9u/2} - 6\pi n^2 e^{5u/2}\right)\exp\left(-\pi n^2 e^{2u}\right)$$

可以看到有 $\Phi(u) = \Phi(-u) \geq 0$，且 $\Phi(u)$ 在 $\mathbb{R}$ 上可积。于是 $\Phi(u)$ 是某种密度函数（差一个常数）。

Conjecture 3.25 (Riemann 猜想)**.** $\xi(s)$ 只有实零点。

可能的一种证明方法是，研究特征函数只有实零点的随机变量。如果可以证明 $\Phi(u)$ 为密度函数的随机变量，是李-杨类随机变量。那么根据之前的讨论，Riemann 猜想就证完了。

结语

路漫漫其修远兮，吾将上下而求索。——离骚