

# 单环马氏链环流的大偏差与涨落定理

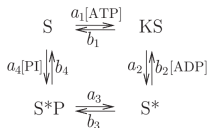
姜瑜浩

指导老师：贾晨教授  
北京计算科学研究中心

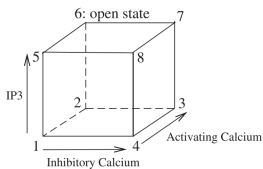
May 15, 2022

# 生化反应

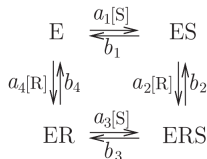
常见的生化反应模型：



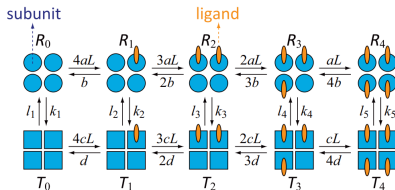
磷酸-脱磷酸化循环



De Young-Keizer 模型



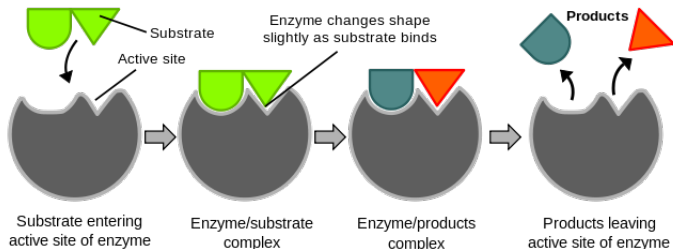
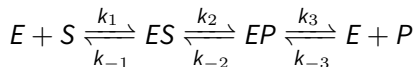
广义的修正机制



MW 模型

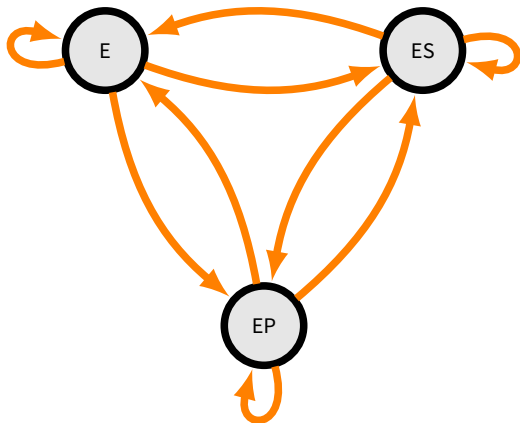
# 酶：生物系统中化学反应的催化剂

三步 Michaelis-Menten 酶动力学模型：



# 酶：生物系统中化学反应的催化剂

## 酶化反应的马氏链模型



# 马氏链中的环流

## 为什么研究上述马氏链中的环？

考虑上述马氏链中，若环  $E \rightarrow EP \rightarrow ES \rightarrow E$  出现的频率大于环  $E \rightarrow ES \rightarrow EP \rightarrow E$  出现的频率，则化学反应整体向正反应进行，否则向负反应进行。通过研究马氏链中每个环出现的频率，可以研究化学反应整体方向以及反应进程。

# 马氏链中的环流

环擦除方式形成的环

## 回路的定义

马氏链的轨迹中，一段路径的起点和终点相同的，且路径中其他状态各不相同，我们称它为回路，如： $i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_s \rightarrow i_1$

## 回路的等价关系

记  $j_1 \rightarrow j_2 \rightarrow \cdots \rightarrow j_r \rightarrow j_1$  为另一个回路，若存在一个整数  $k$  且  $r = s$ ，使得  $j_1 = i_{k+1}, j_2 = i_{k+2}, \cdots, j_r = i_{k+s}$  成立，则称两个回路等价，其中指标  $k+1, k+2, \cdots, k+s$  为模  $n$  后的结果。

# 马氏链中的环流

## 环擦除方式形成的环

### LE 环的定义

依据回路的等价关系，可以定义等价类，我们称这个等价类为环，并且称所有环构成的集合为环空间  $\mathcal{C}$ 。例如，回路  $i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_s \rightarrow i_1$  所属的环为  $c = (i_1, i_2, \cdots, i_s) \in \mathcal{C}$ 。

### LE(环擦除) 环的形成过程

$n$	0	1	2	3	4	5	6	7	8
$\xi_n$	1	2	3	3	2	3	4	1	4
$\tilde{\xi}_n$	[1]	[1,2]	[1,2,3]	[1,2,3]	[1,2]	[1,2,3]	[1,2,3,4]	[1]	[1,4]
形成的环				(3)	(2,3)			(1,2,3,4)	

Table: 导出链的变化过程和轨道中形成的环

# 马氏链中的环流

## 生成树方式形成的环

### 生成树定义

记  $T$  为转移图  $G$  的有向子图，即  $T$  的所有边也是  $G$  的边， $\bar{T}$  表示与  $T$  对应的无向图。满足下列三个条件的  $T$  被称为图  $G$  的生成树（极大树）：

- $T$  是  $G$  的覆盖子图，即  $T$  包含  $G$  的所有顶点。
- $\bar{T}$  是连通的。
- $\bar{T}$  没有回路，其中无向图的回路是顶点到自身的无向路径。

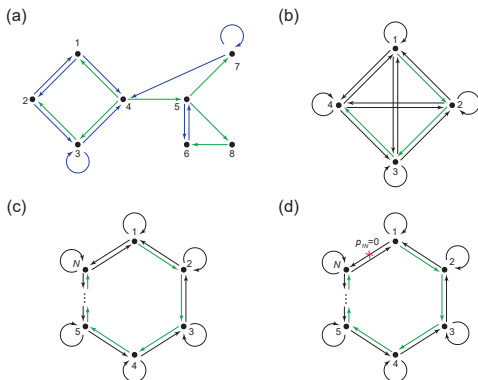
### ST(生成树) 环

一般称有向边  $l \notin T$  为  $T$  的弦。记  $|E| = M$ ,  $|T| = N - 1$  ( $T$  边的数量)，那么任意生成树  $T$  都有  $M - N + 1$  个弦。由于  $\bar{T}$  是连通的且没有回路，如果添加一根弦  $l$  到  $T$ ，会导致无向子图  $\bar{T} \cup \{l\}$  恰好有一条回路。因此  $c_l$  是由此产生的环，并且和  $l$  保持同样的指向。由弦生成的环的集合  $\mathcal{L} = \{c_l : l \in E \setminus T\}$  被称为基本集。



# 马氏链中的环流

## 两种类型环的比较



**Figure:** 马氏链的转移图和相应的生成树 (a) 一般转移图的马氏链，绿色线表示根节点为4的生成树  $T$ ，并且蓝色线表示  $T$  的弦 (b) 四状态全连接马氏链，其中每个状态可以转移到自己和其他状态 (c)  $N$  状态单环马氏链。系统有一个环拓扑。每个状态转移到自身和它两个相邻节点 (d) 一个  $N$  状态的单环马尔科夫链，该系统无法从状态 1 转移到状态  $N$  (b)-(d) 中，绿色箭头表示生成树  $T$

# 马氏链中的环流

## 两种定义方式下的环流

### LE 环流

记  $N_n^c$  为环  $c$  在时刻  $n$  时总共形成的次数，环  $c$  在  $n$  时刻的经验环流为：

$$J_n^c = \frac{1}{n} N_n^c$$

### ST 环流

记  $Q_n^{c_l}$  表示单位时间通过弦  $l$  的次数，ST 弦和弦是一一对应关系，因此可以用通过弦  $l$  的次数定义  $c_l$  形成的次数。 $c_l$  在  $n$  时刻的经验环流为：

$$Q_n^{c_l} = \frac{1}{n} \sum_{i=1}^n 1_{\{\langle \xi_{i-1}, \xi_i \rangle = l\}}$$

# 大偏差理论

## 大数定律

设独立同分布的随机变量序列  $(X_i)_{i \in \mathbb{N}}$  定义于状态空间  $\Gamma = \{1, \dots, r\}$ , 概率分布为  $\rho = (\rho_s)_{s \in \Gamma}$ . 定义  $\rho$  所处的空间为:

$$\mathfrak{M}_1(\Gamma) = \{\nu = (\nu_1, \nu_2, \dots, \nu_r) \in [0, 1]^r : \sum_{s=1}^r \nu_s = 1\}$$

令经验测度  $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ , 即  $L_n \in \mathfrak{M}_1(\Gamma)$ , 则根据大数定律, 有

$$d(L_n, \rho) \xrightarrow{n \rightarrow \infty} 0 \quad \mathbb{P} - a.s.$$

其中上述度量为全变差距离

$$d(\mu, \nu) = \frac{1}{2} \sum_{s=1}^r |\mu_s - \nu_s|, \quad \mu, \nu \in \mathfrak{M}_1(\Gamma)$$

# 大偏差理论

## Level-1 大偏差 (Cramers 定理)

$(X_i)_{i \in \mathbb{N}}$  是上述定义的独立同分布的随机变量序列, 相应的经验测度为  $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ 。那么, 对于  $\forall a > 0$ , 有:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in B_a^c(\rho)) = - \inf_{\nu \in B_a^c(\rho)} I(\nu)$$

其中  $B_a(\rho) = \{\nu \in \mathfrak{M}_1(\Gamma) : d(\nu, \rho) \leq a\}$  是  $\rho$  的邻域,  $B_a^c(\rho) = \mathfrak{M}_1(\Gamma) \setminus B_a(\rho)$ , 且称

$$I_\rho(\nu) = \sum_{s=1}^r \nu_s \log \left( \frac{\nu_s}{\rho_s} \right)$$

为速率函数。

# 大偏差理论

## Level-2 大偏差 (Sanov 定理)

在随机变量序列  $(X_i)_{i \in \mathbb{N}}$  上, 定义对经验测度  $L_n^2 = \frac{1}{n} \sum_{i=1}^n \delta_{(X_i, X_{i+1})}$  不妨令  $X_{n+1} = X_1$  (当  $n \rightarrow \infty$  时, 该假设对  $L_n^2$  取值无影响。) 则  $L_n^2$  所在的概率测度空间  $\widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma) = \{\nu = (\nu_{st}) \in \mathfrak{M}_1(\Gamma \times \Gamma) :$

$$\sum_t \nu_{st} = \sum_t \nu_{ts}, \forall s\}.$$

Sanov 证明了对于  $\forall a > 0$ , 有:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n^2 \in B_a^c(\rho \times \rho)) = - \inf_{\nu \in B_a^c(\rho \times \rho)} I_\rho^2(\nu)$$

其中  $B_a(\rho \times \rho) = \{\nu \in \widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma) : d(\nu, \rho \times \rho) \leq a\}$ ,  $\bar{\nu}_s = \sum_t \nu_{st}$ , 且速率函数为

$$I_\rho^2(\nu) = \sum_{s,t} \nu_{st} \log \left( \frac{\nu_{st}}{\bar{\nu}_s \rho_t} \right)$$

# 环流的大偏差理论

## 环流的大偏差理论

考虑上述环流问题的随机过程  $(X_i)_{i \in \mathbb{N}}$ , 则  $C_\infty = \{c_1, c_2, \dots, c_s\}$  为该过程所有可能环的集合,  $w = (w_{c_1}, w_{c_2}, \dots, w_{c_s})$  为该马氏链的环流分布。不妨令  $X_{n+1} = X_1$ , 环流的经验测度为  $w_n = (w_{c_1, n}, w_{c_2, n}, \dots, w_{c_s, n})$  定义  $w_{c, n}$  所在的空间:

$$\mathfrak{M}_2 = \{\mu = (\mu_{c_1}, \mu_{c_2}, \dots, \mu_{c_s}) \in [0, 1]^s : \sum_{i=1}^s |c_i| \mu_{c_i} = 1\},$$

其中  $|c_i|$  表示环  $c_i$  中点的数量。则对  $\forall a > 0$ , 有:

$$\frac{1}{n} \log \mathbb{P}(w_n \in B_a^c(w)) = - \inf_{\nu \in B_a^c(w)} I^{c_1, c_2, \dots, c_s}(\nu)$$

# 环流的大偏差原理

## 环流大偏差原理的定义

若满足下列三个条件, 则称经验 LE 环流  $(J_n^c)_{c \in \mathcal{C}}$  为满足速率函数为  $I_J: \mathcal{V} \rightarrow [0, \infty]$  的大偏差原理:

- 对于  $\forall \alpha \geq 0$ , 水平集  $\{x \in \mathcal{V} : I_J(x) \leq \alpha\}$  是紧的。
- 对于任意开集  $U \subset \mathcal{V}$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((J_n^c)_{c \in \mathcal{C}} \in U) \geq - \inf_{x \in U} I_J(x). \quad (1)$$

- 对于任意闭集  $F \subset \mathcal{V}$ ,

$$\overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}((J_n^c)_{c \in \mathcal{C}} \in F) \leq - \inf_{x \in F} I_J(x). \quad (2)$$

# 环流的大偏差原理

## 环流大偏差速率函数

定义中的条件 (ii) 和 (iii) 表明, 对任意  $(\nu^c)_{c \in \mathcal{C}} \in \mathcal{V}$ , 速率函数满足:

$$\mathbb{P}(J_n^c = \nu^c, \forall c \in \mathcal{C}) \propto e^{-nI_J(\nu)}, \quad n \rightarrow \infty,$$

假设轨道满足周期边界条件, 还可得 LE 经验环流的联合分布为:

$$\begin{aligned} \mathbb{P}(J_n^c = \nu^c, \forall c \in \mathcal{C}) &= \mathbb{P}(N_n^c = k^c, \forall c \in \mathcal{C}) \\ &= |G_n(k)| \prod_{c \in \mathcal{C}} (\gamma^c)^{k^c}, \end{aligned}$$

其中  $\nu^c = k^c/n$ ,  $G_n(k)$  表示  $n$  时刻可能形成的轨道的集合, 并且称该类轨道为容许轨道。



# 环流的大偏差原理

## 环流大偏差速率函数

例：三状态马氏链的一段轨道中，环的数量  $k = (k^c)_{c \in C}$  为：

$$k^3 = k^{12} = k^{23} = k^- = 1, \quad k^1 = k^2 = k^{13} = k^+ = 0$$

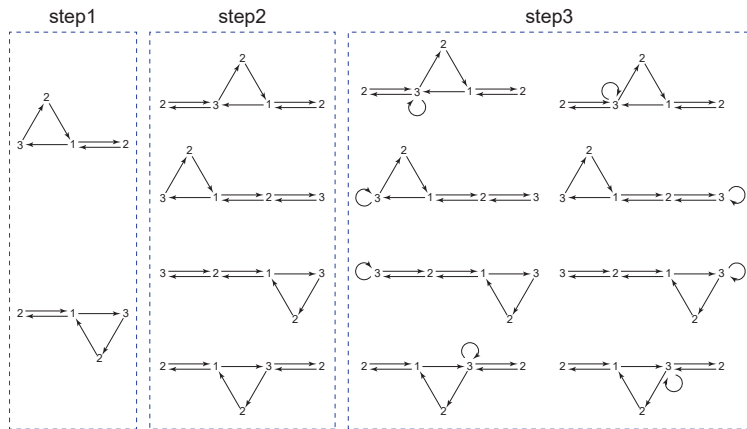
那么在时刻  $n = 8$ ，有 8 个容许轨道，如表2所示。

$m$	0	1	2	3	4	5	6	7	8
$\xi_m$	1	3	3	2	3	2	1	2	1
$\xi_m$	1	3	2	3	3	2	1	2	1
$\xi_m$	1	3	3	2	1	2	3	2	1
$\xi_m$	1	3	2	1	2	3	3	2	1
$\xi_m$	1	2	3	3	2	1	3	2	1
$\xi_m$	1	2	3	2	1	3	3	2	1
$\xi_m$	1	2	1	3	3	2	3	2	1
$\xi_m$	1	2	1	3	2	3	3	2	1

**Table:** 三状态马氏链中，8 个容许轨道，环 (3), (12), (23) 和 (1, 3, 2) 形成一次，环 (1), (2), (13) 和 (1, 2, 3) 没有形成过

# 环流的大偏差原理

## 环流大偏差速率函数



**Figure:** 构建所有容允许轨道的环插入法示意图。依然使用17中的例子。环插入法分为三步：首先我们将所有包含初始状态的环插入轨道，接下来我们将所有剩余的两元环插入轨道，最后我们将所有剩余的一元环插入轨道。经过这三步的环插入，找到了所有八个容许轨道。

# 环流的大偏差原理

## 环流大偏差速率函数

### 计算容许轨道空间的规模

依据图中所示的计算方法，对于所有单环马氏链，可得容许轨道空间的规模为：

$$|G_n(k)| = A_1 A_2 A_3$$

其中：

$$A_1 = \binom{k^1 + k^{12} + k^{1N} + k^+ + k^-}{k^1, k^{12}, k^{1N}, k^+, k^-} := \frac{(k^1 + k^{12} + k^{1N} + k^+ + k^-)!}{k^1! k^{12}! k^{1N}! k^+! k^-!},$$

$$A_2 = \sum_{l^2 + m^2 = k^{23}} \cdots \sum_{l^{N-1} + m^{N-1} = k^{N-1, N}} \prod_{i=2}^{N-1} \binom{l^i + l^{i-1} + k^+ - 1}{l^i} \prod_{i=2}^{N-1} \binom{m^i + m^{i+1} + k^- - 1}{m^i},$$

$$A_3 = \prod_{i=2}^N \binom{\sum_{c \ni i} k^c - 1}{k^i}.$$

# 环流的大偏差原理

## 环流大偏差速率函数

环流大偏差速率函数表达式:

$$\begin{aligned} I_J(\nu) &= -\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(J_n^c = \nu^c, \forall c \in \mathcal{C}) \\ &= -\lim_{n \rightarrow \infty} \frac{1}{n} \left[ \log A_1 + \log A_2 + \log A_3 + \sum_{c \in \mathcal{C}} k^c \log \gamma^c \right]. \end{aligned}$$

化简可得:

$$\begin{aligned} I_J(\nu) &= [h(\nu^{12}) + h(\nu^{1N}) + h(\nu^+) + h(\nu^-) - h(\nu^{12} + \nu^{1N} + \nu^+ + \nu^-)] \\ &\quad + \inf_{X \in V(\nu)} F_\nu(X) + \sum_{i \in S} [h(\nu_i - \nu^i) + h(\nu^i) - h(\nu_i)] - \sum_{c \in \mathcal{C}} \nu^c \log \gamma^c, \end{aligned}$$

其中  $h(x) = x \log x$ ,  $\nu^c = \frac{k^c}{n}$ ,  $\nu_i = \sum_{c \ni i} \nu^c$ .

# 环流的大偏差原理

## 环流大偏差速率函数

由环流大偏差速率函数表达式：

$$\begin{aligned} I_J(\nu) &= -\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(J_n^c = \nu^c, \forall c \in \mathcal{C}) \\ &= -\lim_{n \rightarrow \infty} \frac{1}{n} \left[ \log A_1 + \log A_2 + \log A_3 + \sum_{c \in \mathcal{C}} k^c \log \gamma^c \right]. \end{aligned}$$

环流大偏差速率函数的存在唯一性，转换为优化问题：

$$\inf_{X \in V(\nu)} F_\nu(X) = F_\nu(x^j, y^j),$$

的存在唯一性。

# 环流的大偏差原理

## 环流大偏差速率函数

定义相应的拉格朗日函数：

$$\mathcal{A}_\nu(X, \lambda) = F_\nu(X) + \sum_{i=2}^{N-1} \lambda_i \left( x^i + y^i - \nu^{i,i+1} \right), \quad \forall \nu \in \mathcal{V}$$

其中  $X = (x^i, y^i)_{2 \leq i \leq N-1} \in V(\nu)$  并且  $\lambda = (\lambda_i)_{2 \leq i \leq N-1} \in \mathbb{R}^{N-2}$   
经过化简得到：

$$\frac{x^i}{x^{i-1} + x^i + \nu^+} \frac{x^i + \nu^+}{x^i + x^{i+1} + \nu^+} = \frac{y^i + \nu^-}{y^{i-1} + y^i + \nu^-} \frac{y^i}{y^i + y^{i+1} + \nu^-} = e^{-\lambda_i},$$
$$x^i + y^i = \nu^{i,i+1}, \quad 2 \leq i \leq N-1. \quad (3)$$

其中  $x^1 = \nu^{12}$ ,  $x^N = 0$ ,  $y^1 = 0$  和  $y^N = \nu^{1N}$ 。  
最终我们证明了该方程 3 的解存在且唯一

# 环流的大偏差原理

## 环流大偏差速率函数

环流大偏差速率函数的性质:

- 有界性
- 凸函数
- 对称性: 上述推导中, 都是假设马氏链从状态 1 出发, 我们证明了从任意状态出发都可以得到此结果
- 连续性: 在  $[0, +\infty)$  上连续

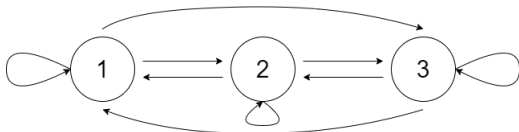
# 环流的大偏差原理

## 环流大偏差速率函数

### 三环马氏链

环流大偏差速率函数，可简化为：

$$I_J(\nu) = \sum_{i \in S} \left[ \nu^i \log \left( \frac{\nu^i / \nu_i}{J^i / J_i} \right) + (\nu_i - \nu^i) \log \left( \frac{(\nu_i - \nu^i) / \nu_i}{(J_i - J^i) / J_i} \right) \right] \\ + \sum_{c \in C, |c| \neq 1} \nu^c \log \left( \frac{\nu^c / \tilde{\nu}}{J^c / \tilde{J}} \right).$$



状态转移图



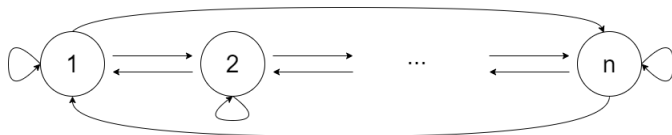
# 环流的大偏差原理

## 环流大偏差速率函数

状态 1 到状态 N 的转移概率为 0 的单环马氏链的速率函数

环流大偏差速率函数，可简化为：

$$I_J(\nu) = \sum_{i \in S} \left[ \nu^i \log \left( \frac{\nu^i / \nu_i}{J^i / J_i} \right) + \nu^{i,i+1} \log \left( \frac{\nu^{i,i+1} / \nu_i}{J^{i,i+1} / J_i} \right) + (\nu^{i-1,i} + \nu^+) \log \left( \frac{(\nu^{i-1,i} + \nu^+) / \nu_i}{(J^{i-1,i} + J^+) / J_i} \right) \right]$$



状态转移图

# 环流的大偏差原理

## 环流大偏差速率函数

### 净环流大偏差速率函数

记 LE 经验净环流为  $(\tilde{J}_n^c)_{c \in \mathcal{C}}$ , 由收缩原理可以得到:

$$\begin{aligned}\mathbb{P}(\tilde{J}_n^+ = x) &= \mathbb{P}(J_n^+ - J_n^- = x) \\ &= \sum_{\nu^+ - \nu^- = x} \mathbb{P}(J_n^c = \nu^c, \forall c \in \mathcal{C}) \\ &\propto \sum_{\nu^+ - \nu^- = x} e^{-nI_J(\nu)}.\end{aligned}$$

因此  $\tilde{J}_n^+$  满足大偏差原理, 相应的速率函数  $I_J$  为:

$$I_J(x) = \inf_{\{\nu \in \mathcal{V}: \nu^+ - \nu^- = x\}} I_J(\nu).$$

# 环流的大偏差原理

一般马氏链的 ST 环流的大偏差

## 回顾对经验测度的大偏差

定义  $n$  时刻的对经验测度  $R_n : E \rightarrow [0, 1]$  为

$$R_n(i, j) = \frac{1}{n} \sum_{m=1}^n 1_{\{\xi_{m-1}=i, \xi_m=j\}},$$

对经验测度满足下面的大偏差原理：

$$\mathbb{P}(R_n(i, j) = R(i, j), \forall \langle i, j \rangle \in E) \propto e^{-n I_{\text{pair}}(R)}, \quad n \rightarrow \infty,$$

其中的速率函数为  $I_{\text{pair}} : \mathcal{M} \rightarrow [0, \infty]$  为

$$I_{\text{pair}}(R) = \sum_{\langle i, j \rangle \in E} R(i, j) \log \frac{R(i, j)}{R(i) p_{ij}}$$

# 环流的大偏差原理

## 一般马氏链的 ST 环流的大偏差

### ST 环流大偏差速率函数

对经验测度可以被分解为下面 ST 环流的加权和：

$$R_n(i, j) = \sum_{c_l \in \mathcal{L}} H^{c_l}(i, j) Q_n^{c_l},$$

其中：

$$H^{c_l}(i, j) = \begin{cases} 1, & \text{if } \langle i, j \rangle \in c_l \text{ and } \langle i, j \rangle \in T, \text{ or } \langle i, j \rangle = l, \\ -1, & \text{if } \langle i, j \rangle \notin c_l, \langle j, i \rangle \in c_l, \text{ and } \langle i, j \rangle \in T, \\ 0, & \text{otherwise.} \end{cases}$$

并且该分解是唯一的。

# 环流的大偏差原理

## 一般马氏链的 ST 环流的大偏差

### ST 环流大偏差速率函数

如果  $R_n = \sum_{c_l \in \mathcal{L}} \nu^{c_l} H^{c_l}$ , 那么对任意  $c_l \in \mathcal{L}$  会有  $\nu^{c_l} = Q_n^{c_l}$ 。由上式分解的唯一性, 可得:

$$\mathbb{P}(Q_n^{c_l} = \mu^{c_l}, \forall c_l \in \mathcal{L}) = \mathbb{P}\left(R_n(i, j) = \sum_{c_l \in \mathcal{L}} \mu^{c_l} H^{c_l}(i, j), \forall \langle i, j \rangle \in E\right) \propto e^{-n I_{\text{pair}}(\sum_{c_l \in \mathcal{L}} \mu^{c_l} H^{c_l})}.$$

这表明 ST 经验环流  $(Q_n^{c_l})_{c_l \in \mathcal{L}}$  满足大偏差原理, 相应的速率函数  $I_Q : \mathcal{M} \rightarrow [0, \infty]$  为:

$$I_Q(\mu) = I_{\text{pair}}\left(\sum_{c_l \in \mathcal{L}} \mu^{c_l} H^{c_l}\right).$$

# 环流的涨落定理

单环马氏链 LE 环流的涨落定理

# 环流的涨落定理

## 单环马氏链 LE 环流的涨落定理

### LE 净环流满足的各种涨落定理

- 暂态涨落定理:

$$\frac{\mathbb{P}(\tilde{J}_n^+ = x)}{\mathbb{P}(\tilde{J}_n^+ = -x)} = \left( \frac{\gamma^+}{\gamma^-} \right)^{nx}.$$

- Kurchan-Lebowitz-Spohn 类型涨落定理:

$$\tilde{g}_n(\lambda) = \tilde{g}_n \left( - \left( \lambda + \log \frac{\gamma^+}{\gamma^-} \right) \right).$$

- 积分涨落定理: 取  $\lambda = -\log \gamma^+ / \gamma^-$  带入上面方程可得

$$\mathbb{E} \left[ e^{\lambda n \tilde{J}_n^+} \right] = 1.$$

- Gallavotti-Cohen 类型涨落定理:

$$\tilde{l}_J(x) = \tilde{l}_J(-x) - \left( \log \frac{\gamma^+}{\gamma^-} \right) x.$$

# 谢谢！