单环马氏链环流的大偏差与涨落定理

姜瑜浩

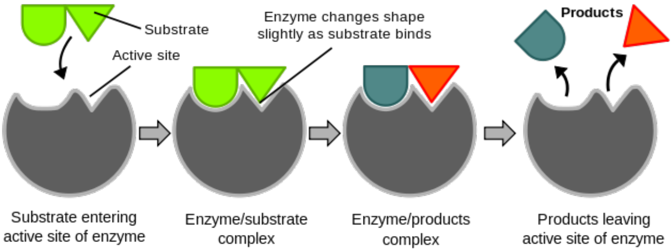
指导老师：贾晨教授
北京计算科学研究中心

May 8, 2022

酶：生物系统中化学反应的催化剂

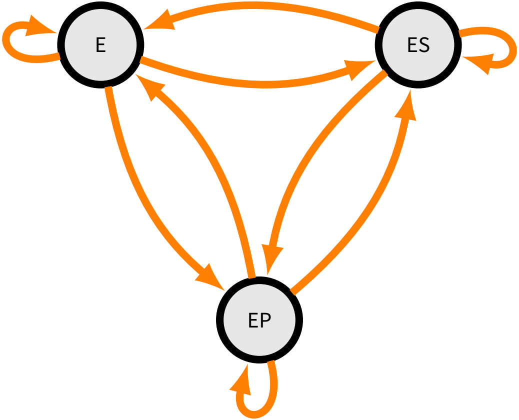
三步 Michaelis-Menten 酶动力学模型：

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EP \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P$$



酶：生物系统中化学反应的催化剂

相应的马氏链模型



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

环擦除定义下的环流

马氏链中的环流

生成树定义下的环流

两种类型环流的比较

环流的大偏差原理

马氏链的环流理论

环流分布

设 $(X_i)_{i\in\mathbb{N}}$ 是平稳不可约正常返马氏链，且转移概率矩阵为 P，状态空间为有限集 Γ。对于该马氏链在 $n>0$ 之前的序列，用 $\mathcal{C}_n$ 表示其中出现的环的集合，$w_{c,n}$ 表示其中环 c 出现的次数。则当 $n\to+\infty$，随机变量序列 $(\mathcal{C}_n, w_{c,n})$ 会有：

$$\mathcal{C}_\infty = \lim_{n\to+\infty} \mathcal{C}_n, \quad a.e.$$
$$w_c = \lim_{n\to+\infty} \frac{w_{c,n}}{n}, \quad a.e.$$

且环流分布为：

$$w_c = p_{i_1,i_2}p_{i_2,i_3}\cdots p_{i_{s-1},i_s}p_{i_s,i_1}\frac{D(\{i_1,i_2,\cdots i_s\}^c)}{\sum_{j\in S} D(\{j\}^c)}$$

其中 $D(H)$ 是以集合 H 为行和列的指标，$I-P$ 的主子式，且令 $D(\phi)=1$。

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 $\Gamma=\{1,\cdots,r\}$，概率分布为 $\rho=(\rho_s)_{s\in\Gamma}$。定义 ρ 所处的空间为：

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

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

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

其中上述度量为全变差距离

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

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：

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

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

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

为速率函数。

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 $n \to \infty$ 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：

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

其中 $B_a(\rho\times\rho) = \{\nu \in \widetilde{\mathfrak{M}}_1(\Gamma\times\Gamma) : d(\nu,\rho\times\rho) \le a\}$，$\bar{\nu}_s = \sum_t \nu_{st}$，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

考虑上述环流问题的随机过程 $(X_i)_{i\in\mathbb{N}}$，则 $C_\infty = \{c_1, c_2, \ldots, c_s\}$ 为该过程所有可能环的集合，$w = (w_{c_1}, w_{c_2}, \ldots, w_{c_s})$ 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 $w_n = (w_{c_1,n}, w_{c_2,n}, \ldots, w_{c_s,n})$ 定义 $w_{c,n}$ 所在的空间：

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

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

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

速率函数

Level-1 大偏差的速率函数

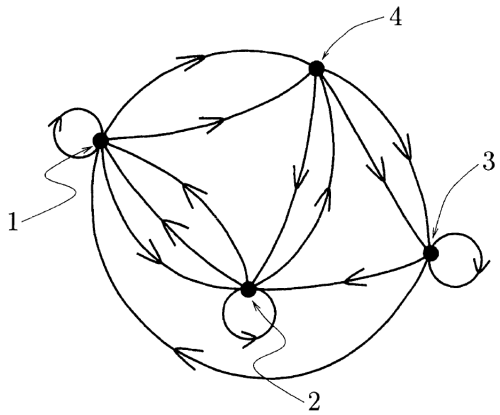
- I_ρ 是有界的，连续的，且在空间 $\mathfrak{M}_1(\Gamma)$ 是强凸的
- $I_\rho \geq 0$。当且仅当 $\nu = \rho$ 时，取等。
- 通常称 $H(\nu|\rho) = I_\rho$ 为 ν 关于 ρ 的交叉熵。在信息论中，通常用其表示两个函数分布之间的差别。

Level-2 大偏差的速率函数

- I_ρ^2 是有界的，连续的，且在空间 $\widetilde{\mathfrak{M}}_1(\Gamma \times \Gamma)$ 是强凸的。
- $I_\rho^2(\nu) \geq 0$。当且仅当 $\nu = \rho \times \rho$ 时，取等。
- $I_\rho^2(\nu) = H(\nu|\bar{\nu} \times \rho)$ 为 ν 相对于 $\bar{\nu} \times \rho$ 的交叉熵。

主要难点

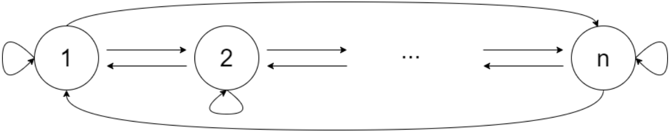
给定一个图 $G^n(k)$，考虑图中遍历所有边的欧拉回路，并计算出每种环出现的数量 $k=(k_c)_{c\in\mathcal{C}_\infty}$ 的欧拉回路数量 $\mathcal{E}(G^n(k))$。



马氏链诱导出的图

难点简化

针对酶动力学模型，下图所示的马氏链



m 状态马氏链诱导出的图

可以得出符合要求的欧拉回路数量：

$$\mathcal{E}_1(G^m(k)) = \binom{k_{12}+k_{1m}+k^{+}+k^{-}}{k_{12},k_{1m},k^{+},k^{-}}\binom{k_1+k_{12}+k_{1m}+k^{+}+k^{-}}{k_{12}+k_{1m}+k^{+}+k^{-}}\left[\Pi_{i\in S,i\neq 1}\binom{\sum_{c\ni i}k_c-1}{\sum_{c\ni i}k_c-k_i-1}\right]$$

$$\left[\sum_{k_{23}^{+}+k_{23}^{-}=k_{23}}\sum_{k_{34}^{+}+k_{34}^{-}=k_{34}}\cdots\sum_{k_{m-1,m}^{+}+k_{m-1,m}^{-}=k_{m-1,m}}\binom{k_{23}^{+}+k_{12}+k^{+}-1}{k_{23}^{+}}\binom{k_{34}^{+}+k_{23}^{+}+k^{+}-1}{k_{34}^{+}}\right.$$

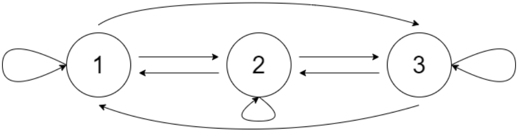
$$\cdots\binom{k_{m-1,m}^{+}+k_{m-2,m-1}^{+}+k^{+}-1}{k_{m-1,m}^{+}}\binom{k_{1,m}+k_{m-1,m}^{-}+k^{-}-1}{k_{m-1,m}^{-}}\binom{k_{m-1,m}^{-}+k_{m-2,m-1}^{-}+k^{-}-1}{k_{m-2,m-1}^{-}}$$

$$\left.\cdots\binom{k_{23}^{-}+k_{34}^{-}+k^{-}-1}{k_{23}^{-}}\right]$$

速率函数

对于三状态马氏链，环流大偏差速率函数：

$$I_3^c(\nu) = \sum_{c\in\mathcal{C}_\infty} \nu_c \log\frac{\nu_c}{w_c} + \sum_{i\in S}(\nu^i - \nu_i)\log\frac{\nu^i-\nu_i}{w^i-w_i}$$
$$- (\tilde{\nu} - \sum_{i\in S}\nu_i)\log(\frac{\tilde{\nu}-\sum_{i\in S}\nu_i}{\tilde{w}-\sum_{i\in I}w_i})$$
$$- \sum_{i\in S}\nu^i\log(\frac{\nu^i}{w^i})$$



状态转移图

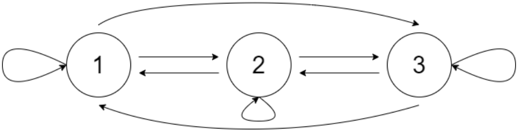
速率函数

对于三状态马氏链，净环流大偏差速率函数：

$$\bar{I}(\bar{\nu}) = \min_{\nu \in \bar{E}(\bar{\nu})} I_3^c(\nu)$$

其中

$\bar{E}(\bar{\nu}) = \{\nu = (\nu_c)_{c \in \mathcal{C}_\infty} \in E | \nu_1 + \nu_2 + \nu_3 + 2(\nu_{12} + \nu_{13} + \nu_{23}) + 3(2\nu_{132} + \bar{\nu}) = 1\}$

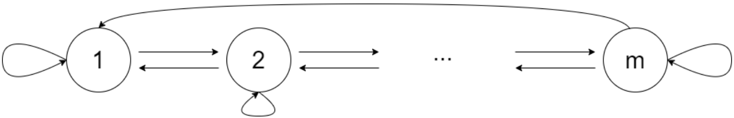


状态转移图

速率函数

对于 n 状态马氏链，且状态间的转移概率非零的仅有 $\{p_{i,i+1}, i = 1, \ldots n-1\} \bigcup \{p_{i,i-1}, i = 2, \ldots n-1\} \bigcup \{p_{n,1}\}$，则相应的大偏差速率函数：

$$I^c_{m'}(\nu) = \sum_{i \in S} \sum_{c \ni i} \nu_c \log \left(\frac{\nu_c}{\nu^i} \Big/ \frac{w_c}{w^i} \right)$$



状态转移图

速率函数

环流大偏差的速率函数

- $I^{c_1,c_2,\ldots,c_s}$ 是有界的，连续的，且是凸函数的。
- $I^{c_1,c_2,\ldots,c_s}(\nu) \geq 0$。当且仅当 $\nu = w$ 时，取等。
- 如果 c_k 和 c_l 相似（环包含的节点相同），则有下式成立（涨落定理）

$$
\begin{aligned}
& I^{c_1,c_2,\ldots,c_s}(x_1,\ldots,x_k,\ldots,x_l,\ldots,x_r) \\
&= I^{c_1,c_2,\ldots,c_s}(x_1,\ldots,x_l,\ldots,x_k,\ldots,x_r) - (\log \frac{\gamma^{c_k}}{\gamma^{c_l}}(x_k - x_l))
\end{aligned}
$$

其中对于 $c = (i_1, i_2, \ldots, i_s)$, 定义 $\gamma^c = p_{i_1 i_2} p_{i_2 i_3} \ldots p_{i_s i_1}$。

谢谢！