Day 6. Large Deviations: Theory & Practice.

Large Deviation In a Nutshell.

大偏差理论旨在描述经验均值偏离其预期值的概率随样本数目的增多而指数下降的行为。我们考虑一个最简单的情形：一个成功概率为 r 的硬币被抛(独立地)，共进行了 T 次的投掷 $(T \gg 1)$。问成功次数 T 除以所有的比例 $f = \tau/T$ 的分布是什么样子的？（我们可以将 τ 称为"广延量"，而将 f 称为"强度量"）。τ 的分布是根据二项分布给出的：

$$P_\tau(T) = C_T^\tau \cdot r^\tau (1-r)^{T-\tau}.$$

注意到它其实是一堆独立随机变量 $x(t)$ 之和 $\tau = \sum_{t=1}^{T} x(t)$ 的分布，所以根据中心极限，在 T 相当大时，这一分布可由正态近似描述：

$$p(f,T) = \sqrt{\frac{T}{2\pi r(1-r)}} \cdot \exp\left(-T \cdot \frac{(f-r)^2}{2r(1-r)}\right).$$

数值计算表明，在 $f \sim r$ 处这一近似是相当好的。但在 f 远远偏离 r 的地方，这样的近似就存在偏差。我们希望对这些稀有事件找个更好的近似。例如，就在我们介绍的例子中，一个远有更好的近似是 $\lim_{T\to+\infty} \frac{1}{T} \ln p(f,T) = -I(f)$，或者 $T \to +\infty$ 时，$p(f,T) \approx \exp(-T \cdot I(f))$。对于二项分布，$I(f)$ 恰为 KL. Div：即 $I(f) = f \cdot \ln \frac{f}{r} + (1-f) \ln \frac{1-f}{1-r}$。服从 $p(f,T) = \exp(-T \cdot I(f))$ 这样的，体现稀有事件发生的概率随着试验次数的上升而增大的关系，都被称为大偏差原理，$I(f)$ 称为 rate function。

一般来说 rate function 有如下性质：1). 非负，且在 $f = \langle f \rangle$ 处 $I(f) = 0$；2). 根据中心极限定理，在 $f \sim \langle f \rangle$ 处通常与抛物线接近（相当于中心极限）。而在 $|f - \langle f \rangle|$ 很大时，$I(f)$ 往往与抛物线差很远。直观上来看，对概率自对数就等于放大了那些原本稀有的事件。所以你可以认为大偏差是一种更精确的中心极限。若在 $I(f)$ 的极小值附近作泰勒展开，则它就会回到中心极限的情形：$I(f) = \frac{(f - \langle f \rangle)^2}{2\tilde{\sigma}_f^2}$，其中 $\tilde{\sigma}_f^2 = \lim_{T\to+\infty} T \cdot \sigma_f^2$。

我们希望给出 $I(f)$ 的具体形式，这需要 cumulant generating function。$\phi^{(f)}(q;T) = \ln \langle \exp(T \cdot q f) \rangle = \ln \langle \exp(q\tau) \rangle$。不难意识到，这个函数会随着 T 线性增长。所以，我们研究所谓 scaled CGF：$\psi^{(f)}(q) = \lim_{T\to+\infty} \frac{1}{T} \cdot \ln \langle \exp(qTf) \rangle$。不难证明从 scaled CGF 可得到随机变量的一些信息。e.g. $\partial_q \psi^{(f)}(q;T)|_{q=0} = \langle f \rangle$，$\partial_q^2 \psi^{(f)}(q;T)|_{q=0} = \tilde{\sigma}_f^2$。

Gärtner — Ellis Theorem 指出：$I(f)$ 与 $\psi^{(f)}(q)$ 由 Legendre 变换有关：$I(f) = \sup_q [f \cdot q - \psi^{(f)}(q)]$。这可以有一个启发式的证明：

$$\psi^{(f)}(q) = \lim_{T\to+\infty} \frac{1}{T} \ln \int_{-\infty}^{+\infty} df \cdot p(f,T) \cdot \exp(T \cdot q \cdot f)$$

$$= \lim_{T\to+\infty} \frac{1}{T} \ln \int_{-\infty}^{+\infty} df \cdot \exp[-T(I(f) - qf)] = \sup_f [qf - I(f)].$$

（只有 $I(f) - qf$ 极小值处有贡献）。

Currents, Traffics and Other Observables.

在之前，我们已介绍了几个主要的可观测量：$w(\vec{x})$，$q(\vec{x})$，$s^{tot}(\vec{x})$。它们都是从轨迹中推导而来的。大偏差可用于研究它们在很长一段时间 T 内的情形。这些是广延量，我们除以 T 便得到强度量。

现在我们研究两类可观测量：一类称为 Static observables，广延的 S_0 可以表示为：

$A(\vec{x}) = \sum_y a_y \tau_y$，其中 τ_y 是系统位于 y 态的时间，$\tau_y(\vec{x}) = \int_{t_0}^{t_f} dt \cdot \delta^{k}_{y, x(t)}$

另一类则称为 Dynamic observables，它与系统在两态之间跳跃次数有关，即 $J(\vec{x}) = \sum_{x \neq x'} n_{xx'}(\vec{x}) \cdot \underline{\underline{d_{xx'}}}$。被称为"距离"。$n_{xx'}(\vec{x}) = \sum_{k=1}^{n} \delta^k_{x x_k} \delta^k_{x' x_{k+1}}$

若 $d_{xx'}$ 是反对称的，则这样的可观测量 $J(\vec{x})$ 被称作 Integrated Empirical Currents。它们在时间反演 $\vec{x} \to$ 下反向。典型的例子如热产生的增 ($d_{xx'} = k_B \cdot \ln\left(\frac{k_{xx'}}{k_{x'x}}\right)$) 或 整个系统的熵增 ($d_{xx'} = k_B \ln \left[\frac{k_{xx'} p_x}{k_{x'x} p_{x'}} \right]$)。

而将停留时间与总时间的比值称为经验频率: $f_y = \tau_y/T = \frac{1}{T}\int_{t_0}^{t_f} dt\, \delta_{y,x(t)}$. 同样你可定义 current 的强度版本 $j = J(x)/T$. $j_{xx'} = \frac{J_{xx'}(x)}{T} = \frac{n_{xx'} - n_{x'x}}{T}$ 代表特定两态间 current.

有时 $d_{xx'}$ 可以是对称的. e.g. 考虑一对态 x, x' 之间 traffic. $t_{xx'} = \frac{1}{T}(n_{xx'}(\bar{x}) + n_{x'x}(\bar{x}))$.

显然, 对于任一处于 steady state 的系统, 它们可观测量的强度版本服从大偏差.

对于在小的 driving 下远离平衡态的系统 (但它处于 steady state) 它的流 j 的速率函数是二次的. 下面给一个证明. 首先考虑处于平衡态的系统, 此时 j 的分布关于 0 对称, 从而有

$\psi^{(j)}(q) = \psi^{(j)}(-q)$. 而存在了 driving 之后, 这样的对称性被打破. 不妨将此时 $\psi(q)$ 视作在偶函数基础上进行了微扰. 从而 $\psi(q) = \psi^+(q) + \psi^-(q)$. 展开 ψ^+, ψ^- 有 $\begin{cases} \psi^+(q) = \psi_2 q^2 + \psi_4 q^4 \\ \psi^-(q) = \psi_1 q + \psi_3 q^3 \end{cases}$

$I(j) = \sup_q [qj - \psi^{(j)}(q)] = q^* j - \psi^*(q^*)$. 代入 $q^* = \frac{1}{2\psi_2}(j - \psi_1\delta) \Rightarrow I(j) = \frac{(j-\psi_1\delta)^2}{4\psi_2}$

对于这类情形 (下一p), 可以定义新的观测量, 例如此时 current 关于 dx 的线性泛函.

→ Large Deviations and Fluctuation Relations.

这节, 我们将大偏差和涨落关系结合起来. 定义一段时间 T 内的总熵产率 $j_S^{tot} = S^{tot}/T$. 从而涨落关系为: $\frac{P(j_S^{tot}, T)}{P(-j_S^{tot}, T)} = \exp(T j_S^{tot}/k_B)$.

而 $P(j_S^{tot}, T)$ 又服从大偏差原理, 从而有: $P(j_S^{tot}, T) \approx \exp(-T\, I(j_S^{tot}))$ 从而有所谓精细的渐近涨落关系. $I(j_S^{tot}) - I(-j_S^{tot}) = -j_S^{tot}/k_B$.

e.g. 特别地, 若 j_S^{tot} 的分布是高斯的, 则上式通过匹配两侧相等可得: $2k_B\langle j_S^{tot}\rangle = \tilde{\sigma}^2_{j_S^{tot}}$ 既然 rate function 和 SCGF 有关系, 则我们也可以将涨落关系以 SCGF 的形式写出.

$\psi^{(j_S)}(q) = \sup_{j_S}[j_S q - I(j_S)] = \sup_{j_S}[j_S(q + \frac{1}{k_B}) - I(-j_S)] = \psi^{(j_S)}(-q - \frac{1}{k_B})$.

→ Fluctuation Theorem for Currents.

前面, 我们说明了一个图上的循环率流可以被分解至若干基本环上. 尤其这种方法可用于研究具有非平衡稳态的系统. 现在我们讨论这一表述和大偏差理论的联系. 我们仍从涨落关系 $P_{\bar{x}} \cdot \exp(-S^{tot}(\bar{x})/k_B) = P_{\tilde{x}}$ 开始.

这里, 我们只考虑全部处于环上的路径. 首先, 想象的熵产生可以被分解到环上. 记第α环的 Affinity. $A_\alpha = k_B T \ln\left(\prod_{j=1}^{n} \frac{k_{x_j x_{j+1}}}{k_{x_{j+1} x_j}}\right)$. 其表示了在图上走了一个"环", 即系统从某态出发, 又回到这一态的过程.

这个过程的熵产生可以被分解到基本环上. $S^{res}(l) = \sum_\alpha c_\alpha(l)\frac{A_\alpha}{T}$. c_α 表示正向/反向. 从而, 一条轨迹上的熵产生为: $S^{res}(\bar{x}) = \sum S^{res}(l) + S^{rem} = \sum_\alpha \frac{A_\alpha}{T} n_\alpha(\bar{x}) + S^{rem}$ "remainder entropy". 即在轨迹路径中多余的熵.

计算 $\psi^{(\vec{J})}(\vec{q}) = \lim_{T\to+\infty} \frac{1}{T}\ln\left\langle \exp\left(\sum_\alpha q_\alpha n_\alpha(\bar{x})\right)\right\rangle$. 在 $T\to\infty$ 时, 未成环的路径上的熵 S^{rem} 和系统的熵变 ΔS^{sys} 都有界, 其相对于环上贡献的部分随时间 T 增长. 从而有:

$$\psi^{(J)}\left(q - \frac{A}{k_BT}\right) = \lim_{T\to+\infty} \frac{1}{T}\ln\int D\bar{x}\, P_{\bar{x}} \exp\left(\sum_\alpha q_\alpha n_\alpha(\bar{x}) - S^{tot}(\bar{x})/k_B\right)$$
$$= \lim_{T\to+\infty} \frac{1}{T}\ln\int D\bar{x}\, P_{\tilde{x}} \exp\left(\sum_\alpha q_\alpha n_\alpha(\bar{x})\right)$$
$$= \lim_{T\to+\infty} \frac{1}{T}\ln\int D\tilde{x}\, P_{\tilde{x}} \exp\left(-\sum_\alpha q_\alpha n_\alpha(\tilde{x})\right)$$
$$= \lim_{T\to+\infty} \frac{1}{T}\ln\int Dx\, P_x \exp\left(-\sum_\alpha q_\alpha n_\alpha(x)\right) = \psi^{(J)}(-q).$$

所以 $\psi^{(J)}(q)$ 有 $q - A/k_BT \longleftrightarrow -q$ 的对称性. 下面我们说明 Onsager 倒易关系中有这种对称性.

平均的沿着α环流动的流就为 $J_\alpha(A) = \langle j_\alpha\rangle = \frac{\partial\psi^{(J)}}{\partial q_\alpha}\Big|_{q=0}$. 这里 $\psi^{(J)}$ 被视为各个环上熵产生 A_α 的函数.

在各个 A_α 都很小的情形下，可以有 $T_\alpha(A) = \sum_\beta L_{\alpha\beta} A_\beta + o(A)$。从而我们有：$L_{\alpha\beta} = \frac{\partial^2 \psi(j)}{\partial q_\alpha \partial A_\beta}\Big|_{q=0,A=0}$。利用上面的对称性，我们有：

$$\frac{\partial^2 \psi(j)}{\partial q_\alpha \partial A_\beta}\Big|_{q,A} = \frac{\partial^2 \psi(j)}{\partial q_\alpha \partial A_\beta}\Big|_{-q-A/k_BT,\,A}$$

→ 这里还要算好符号

$$= \underbrace{\frac{1}{k_BT}\frac{\partial^2 \psi(j)}{\partial q_\alpha \partial q_\beta}\Big|_{\vec{q}=-\vec{q}}}-\vec{A}/k_BT - \frac{\partial^2 \psi(j)}{\partial q_\alpha \partial A_\beta}\Big|_{-q-A/k_BT,\,A}$$

取 $q=A=0$，我们得到了"详细的倒易关系"：$L_{\alpha\beta} = \frac{1}{2k_BT}\cdot\frac{\partial^2 \psi(j)}{\partial q_\alpha \partial q_\beta}\Big|_{q,A=0} = L_{\beta\alpha}$

→ Tilting

由大偏差可以看到，一个可观测量的 Scaled CGF 有非常重要的信息，但很多时候 SCGF 很难直接计算。此时的计算依赖于一种称作 Tilting 的技术。首先定义条件生成函数：

$\phi_x^{(j)}(q,t) = \langle \exp(q\cdot T\cdot j) \mid x(t)=x \rangle = \int D\tilde{x}\cdot P\tilde{x}\cdot \delta^K_{x,x(t)} \exp(q\cdot j[\tilde{x}]T)$ 原本的 CGF：$\psi^{(j)}(q,t) = \sum_x \phi_x^{(j)}(q,t)$。我们显式地展开 P，有：

$$\phi_x^{(j)}(q,t) = \int D\tilde{x}\cdot \delta^K_{x,x_n} \exp\Big(q\cdot \sum_{k=1} d_{x_k,x_{k-1}}\Big)\, k_{x_n x_{n-1}}\cdot \exp(-k^{out}_{x_n}\cdot(t-t_{n-1}))\cdots \exp(-k^{out}_{x_1}(t_2-t_1))\, k_{x_1x_0}\cdot \exp(-k^{out}_{x_0}(t_1-t_0))\, p_{x_0}(t_0)$$

$$= \int D\tilde{x}\cdot \delta^K_{x,x_n} \exp(-k^{out}_{x_n}(t_n-t_{n-1}))\,(k_{x_nx_{n-1}} \exp(q\cdot d_{x_nx_{n-1}}))\cdots(k_{x_1x_0}\exp(q\cdot d_{x_1x_0}))\exp(-k^{out}_{x_0}(t_1-t_0))\, p_{x_0}(t_0).$$

所以，从这里看，$\phi_x^{(j)}(q,t)$ 可以被写成一个路径积分的形式，只不过这个路径是由一个"倾斜"主方程给出的：

$\frac{d}{dt}\phi_x^{(j)}(q,t) = \sum_{x'} L_{xx'}(q)\cdot \phi_{x'}^{(j)}(q,t)$ 其中，$L_{xx'}$ 为倾斜生成元。$L_{xx'}(q) = \begin{cases} k_{xx'}\exp(q\cdot d_{xx'}) & x\neq x' \\ -k_x^{out} & x=x' \end{cases}$ 显然，$L_{xx'}$ 并不一定满足 $\sum_x L_{xx'}(q)=0$，从而 $\phi_x^{(j)}(q,t)$ 并非归一化的。

在 $L_{xx'}$ 时不变的情形下，倾斜主方程的解为：$\phi_x^{(j)}(q,t) = \sum_{x'}\big[\exp((t-t_0)\cdot L(q))\big]_{xx'} p_{x'}(t_0)$。从而 $\phi_x^{(j)}(q,t)$ 的演化是被矩阵 $\{L_{xx'}\}$ 的最大本征值控制的，从而我们可以估计：

$\phi_x^{(j)}(q,t) \approx \exp((t-t_0)\cdot\lambda_{max}(q)) \Rightarrow \psi^{(j)}(q) = \lambda_{max}(q)$。从而，只需计算倾斜主方程生成元的本征值，就可以计算 $\psi^{(j)}(q)$。对于状态数很少的情况，本征值可通过对 $L_{xx'}$ 直接对角化计算，而在状态数很多时，还有更简单的计算方法：矩阵 L 的本征方程为 $\det(L(q)-\lambda I) = \sum_{n=0}^{N} a_n(q)\cdot\lambda^n = 0$。由于对于 $q=0$ 时倾斜的 L 回归到未倾斜的情形，所以，必然有 $\lambda_{max}(q=0)=0$。

考虑函数 $g(\lambda,q) = \sum_{n=0}^{N} a_n(q)\cdot\lambda^n(q) = 0$，它对所有 q 恒成立，从而有：

$$\begin{cases} \frac{d}{dq} g(\lambda(q),q) = \frac{\partial g}{\partial\lambda}\cdot\frac{\partial\lambda}{\partial q} + \frac{\partial g}{\partial q} = 0 \\ \frac{d^2}{dq^2} g(\lambda(q),q) = \frac{\partial^2 g}{\partial\lambda^2}\cdot\left(\frac{\partial\lambda}{\partial q}\right)^2 + \frac{\partial g}{\partial\lambda}\cdot\frac{\partial^2\lambda}{\partial q^2} + 2\frac{\partial^2 g}{\partial\lambda\partial q}\cdot\frac{\partial\lambda}{\partial q} + \frac{\partial^2 g}{\partial q^2} = 0 \end{cases}$$

利用它们，我们可以算出随机变量 j 的一阶、二阶信息：

$\bar{j} = \partial\lambda/\partial q\,|_{q=\lambda=0}$；$\tilde{\sigma}_j^2 = \partial^2\lambda/\partial q^2\,|_{\lambda=q=0}$

另外一个方法是使用数值方法计算 $\lambda_{max}(q)$。定义 $\Sigma_{xx'} = \delta^K_{xx'} + \Delta t\cdot L_{xx'}$，$\Delta t$ 是一个很小的正数。从而 $\Sigma_{xx'}$ 没有负元素，则 $\Sigma_{xx'}$ 的最大本征值为 $1+\Delta t\cdot\lambda_{max}(q) = \tilde{\lambda}_{max}(q)$

将 Σ 作用在任意的初始向量 $\phi_x^{(0)}$ 上，从而，对于很大的 n，我们有：$\Sigma^n\cdot\phi^{(0)} \sim (\tilde{\lambda}_{max})^n\cdot\phi_{max}$.

→ Example: Michaelis-Menten Reaction Scheme.

我们考虑一个经典的化学反应模型：$A+E \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} E^* \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} B+E$，$E$ 是酶，E^* 是与反应物结合的酶。这里考虑最简单的情形，各个反应物的浓度保持不变且酶只有一个。

该跳跃满足 $\frac{k_1^+k_2^+}{k_1^-k_2^-} = \exp\left(\frac{1}{k_BT}(-\varepsilon+\mu_A+\mu_B)\right) = \exp\left(-\frac{1}{k_BT}\Delta G\right)$，$\Delta G$ 是一个B分子与一个A分子之间的自由能差。为了写出主方程，我们设置 $\begin{cases} k_1^+ = \omega_1\exp\left((2\mu_A-\varepsilon^*)/2k_BT\right);\ k_1^- = \omega_1\exp\left(\varepsilon^*/2k_BT\right). \\ k_2^+ = \omega_2\exp\left(-(\varepsilon-\varepsilon^*)/2k_BT\right);\ k_2^- = \omega_2\exp\left((2\mu_B+\varepsilon-\varepsilon^*)/2k_BT\right). \end{cases}$

并假设第二个反应远慢于第一个，即 $\omega_2 \ll \omega_1$。我们用两个变量描述这个系统，一个是目前酶有没有被结合的0/1变量，另一个是自反应开始以来生成的B分子数目n（所以它可以为负）。相应主方程为：

$$\begin{cases} \frac{dp_{0n}}{dt} = \underset{\text{被结合的脱离}}{k_1^- p_{1n}} + \underset{\text{从其被结合生成一个}}{k_2^+ p_{1,n-1}} - (k_1^+ + k_2^-)p_{0n} \\ \frac{dp_{1n}}{dt} = k_1^+ p_{0n} + k_2^- p_{0,n+1} - (k_1^- + k_2^+)p_{1n} \end{cases}$$

记酶被结合的概率 $p=\sum_n p_{1n}$，从而有：$\frac{dp}{dt} = \sum_{n=-\infty}^{+\infty}\left(k_1^+ p_{0n} + k_2^- p_{0,n+1} - (k_1^- + k_2^+)p_{1n}\right)$，$\Rightarrow dp/dt = (k_1^+ + k_2^-)(1-p) - (k_1^- + k_2^+)p$。

可解出其稳态 $p^{st} = \left(1+\frac{k_1^-+k_2^+}{k_1^++k_2^-}\right)^{-1} \underset{(\omega_2\ll\omega_1)}{\approx} \frac{\exp\left(\frac{1}{k_BT}(\mu_A-\varepsilon^*)\right)}{1+\exp\left(\frac{1}{k_BT}(\mu_A-\varepsilon^*)\right)}$。此时，产物B的产生率为：$\langle \dot n_B\rangle = k_2^+ p^{st} - k_2^-$ → $p_{1n}\to p_{1,n+1}$ 的总速率为 k_2^-，$p_{0n}\to p_{0,n+1}$

当酶在结合态的时候，以 p^{st} 的概率生成产物

现在我们来应用大偏差理论，首先定义CGF。我们对于不同状态下酶结合的0/1变量取条件来得到条件CGF。具体而言，取 $\phi_a(q,t) = \sum_{n=-\infty}^{+\infty}\exp\left(q(n+a/2)\right)\cdot p_{a,n}(t)$。

你可以把a这个系数视作一个有两个状态的新current，从而将CGF作为均值用路径测度展开，但这里直接从master 开始。下面可以求出：

$$\frac{d\left(\sum_n \exp(q(n+0/2))\cdot p_{0n}\right)}{dt} = k_1^- \sum_n\left(\exp(q(n+1/2))\exp(-q/2)\, p_{1n}\right) + k_2^+\sum_n\left(\exp(q((n-1)+1/2))\cdot\exp(q/2)\cdot p_{1n}\right)$$

另一条方程同理，从而我们可以得出 $\phi_a(q,t)$ 的演化：

$$\frac{d}{dt}\begin{pmatrix}\phi_0\\ \phi_1\end{pmatrix} = L(z)\begin{pmatrix}\phi_0\\ \phi_1\end{pmatrix},\quad L(z) = \begin{pmatrix} -(k_1^++k_2^-) & z^{-1}k_1^- + z k_2^+ \\ z k_1^+ + z^{-1} k_2^- & -(k_1^-+k_2^+)\end{pmatrix},\quad z=\exp(q/2).$$

下一步是给出 scaled CGF，从而要算L的本征值。不难发现，对于变换 $z'^2 = \frac{k_1^-k_2^-}{k_1^+k_2^+z^2}$（或 $q\to q' = \frac{\Delta G}{k_BT}-q$），$L(z)$ 的本征值不变。下面求n的CGF：

$\phi^{(n)}(q,t) = \sum_n \exp(qn)\left[p_{0,n}(t)+p_{1,n}(t)\right] = \sum_n \exp(qn)\, p_n(t) = \phi_0(q) + \exp(-q/2)\phi_1(q) \asymp \exp(t\lambda_{max}(q))$。定义速率 $\dot n = n/t$。不难给出 $\psi^{(\dot n)}(q) = \lambda_{max}(q)$，并由对称性得到 $\lambda_{max}(q)=\lambda_{max}\left(\frac{\Delta G}{k_BT}-q\right)$。

利用 $\lambda_{max}(q)$ 的解析表达式或近似表达式 $\lambda_{max}(q) = \begin{cases}\sqrt{k_1^+k_2^+}\cdot\exp(q/2) & q\to+\infty \\ \sqrt{k_1^-k_2^-}\cdot\exp(-q/2) & q\to-\infty\end{cases}$，就可以计算所有统计量，例如 $\langle\dot n\rangle = d\lambda_{max}/dq\,|_{q=0}$，以及在 $\dot n$ 很大时，$I(\dot n) = 2|\dot n|\left(\ln\frac{2|\dot n|}{\sqrt{k_1^+k_2^+}}-1\right)$。

➜ Example 2: Fluctuation Relations in a model of kinesin

➜ Cloning

在计算机上模拟一个主方程的轨迹是容易的，但研究大偏差事件却是不容易的——不断采样，你会不断延长采样时间以期让这些稀有事件出现。但根据大偏差，这些稀有事件发生的概率会随T的上升而指数级地减少，从而你永远等不到这样的稀有事件。若从另外一条路走，则计算倾斜的生成元 $L_{xx'}$ 的最大本征值也并非易事。所以可以尝试直接模拟 $\phi_x(q,t)$ 满足的"倾斜主方程"。对于概率密度守恒的主方程，模拟是简单的：只需生成指数分布的等待时间，并依概率在状态间跳跃即可。但 $\phi_x(q,t)$ 满足的主方程的概率密度并不守恒，所以如何对其进行模拟是个问题。一种用于解决此问题的技术称为 cloning，其做法有点类似于 [illegible] —— 我们同时模拟N个系统。为此，我们先做如下定义：

- 倾斜跳率 $k_{xx'}(q) = k_{xx'}\exp(q\, d_{xx'})$。 • 倾斜脱离率 $k_x^{out}(q) = \sum_{x'\neq x} k_{x'x}(q)$。
- 倾斜等待时间分布：$\rho_x(t;q) = k_x^{out}(q)\exp(-k_x^{out}(q)t)$。 • 权重/weights：$Y_x(t;q) = \exp\left[(k_x^{out}(q) - k_x^{out})t\right]$。

利用这些重新写 $\phi_x^{(\dot n)}(q,t) = \int Dx\,\delta^K_{x,x_n}\frac{1}{k_{x_n}^{out}(q)}\rho_{x_n}(t-t_n;q)\cdot Y_{x_n}(t-t_n;q)\cdot\frac{k_{x_nx_{n-1}}(q)}{k_{x_{n-1}}^{out}(q)}\cdot\rho_{x_{n-1}}(t_n-t_{n-1};q)\cdot Y_{x_{n-1}}(t_n-t_{n-1};q)\cdots\frac{k_{x_2x_1}(q)}{k_{x_1}^{out}(q)}\cdot\rho_{x_1}(t_2-t_1;q)\cdot Y_{x_1}(t_2-t_1;q)$

$\cdot\frac{k_{x_1x_0}(q)}{k_{x_0}^{out}(q)}\rho_{x_0}(t_1;q)\cdot Y_{x_0}(t_1)\cdot p_{x_0}(t_0)$。

我们需要对它进行数值模拟。为此，我们要选取系统的 N 份 clone，$N \gg 1$。在时刻 t 时，第 α 个系统的状态由 x^α 表征，它下一次跳跃的时刻由 t^α 表征。

在初始时，每一个系统的状态 $x^\alpha(t=0)$ 是从初始分布 $p_x(t_0)$ 中随机抽取的。而跳跃之间的间隔 Δt_α 也是从分布 $p_{x_\alpha}(\Delta t, q)$ 中抽取的。在初始采样后，设置 $t^\alpha = t_0 + \Delta t_\alpha$。之后进行如下迭代：

- 找到最先要进行跳跃的系统编号 α^*，即 t^{α^*} 为所有 t^α 中最小者
- 将 t 移动到 t^{α^*}。
- 更新 x^{α^*} 的状态 x。新状态从跳转概率 $p_{x'|x^{\alpha^*}} = k_{x'x^{\alpha^*}}(q) / k^{out}_{x^{\alpha^*}}(q)$ 中随机抽取的。
- 将 α^* 系统下一跳的时间设置到 $t^{\alpha^*} = t + \Delta t$。$\Delta t$ 是从 $p_x(\Delta t, q)$ 中随机抽取的。
- 计算这一段时间的权重 $Y_x(\Delta t) = \exp[(k_x^{out}(q) - k_x^{out})\Delta t]$
- 根据权重调整 clone。具体而言，应计算 $y = \lfloor Y_x(\Delta t) + \varepsilon \rfloor$，$\varepsilon \sim U[0,1]$
  - 若 $y=0$，则直接将这个 clone 删除。
  - 若 $y=1$，不进行任何操作。
  - 若 $y>1$，则把这个系统复制 $y-1$ 份。
- 将 clone 的数目调整至 N
- 保存一个 rescaling factor，$X = N/(N+y-1)$.

以上算法的正确性证明可能并非那么显然的，但我们得到的结论是：你可以认为 $Y_x(\Delta t)$ 在 Δt 时间内，路径对于样本事件出现的贡献。

对于 CGF 的结果通过：$p_x^{(s)}(q,t) = (X_1 X_2 \cdots X_n)^{-1}$. 从而 scaled CGF 为：$\psi(q) = -\lim_{t\to\infty} \frac{1}{t} \ln(X_1 \cdots X_n)$.

→ Levels of Large Deviations

在前面，我们利用大偏差原理研究了随机可观测量（static 和 dynamic）。但有些时候我们希望处理的是某轨迹的泛函。这时，我们需要引入不同层级 (level) 上的大偏差原理。

- level 1 是对于瞬时状态的函数的大偏差原理 e.g. 系统的能量 $E_{x(t)}$
- level 2 描述一些静态可观测量（与系统在各态上的时间有关）。主要以 empirical vector 为代表：$f = (f_x)$，f_x 为 x 态上的停留时间占比。
- level 3 则是轨迹的泛函，即 $\omega = (x_t, x_{t+1}, \cdots, x_{t+n})$ 上的可观测量。

为了处理不同层级上的大偏差原理，我们需要引入所谓 Contraction principle / 收缩原理：若我们已知随机变量 x 的 rate function 为 $I^{(x)}(x)$，x, y 间有函数关系 $y = f(x)$，则 y 的 rate function 为：

$$I^{(y)}(y) = \inf_{x: f(x)=y} I^{(x)}(x)$$

直观上说，y 的速率函数 $I^{(y)}(y)$ 在某点的值由 x 对应的"最可能事件"决定。

对于马尔可夫系统，其实可以定义一个"Level 2.5"，即 empirical vector $f=(f_x)$ 与 empirical jump frequency $r_{xx'} = \frac{n_{xx'}(\omega)}{T}$ 的联合分布。对于由主方程控制的系统，Level 2.5 很重要，因为一些重要可观测量 e.g. 熵产生，的大偏差原理可以从中导出。因此我们将推导 Level 2.5 的 rate function.

这种系统处于稳态，并在时间段 $[0,T]$，为简单起见，我们假设 empirical stationary 假设：$\sum_x r_{xx'} = \sum_{x'} r_{x'x}$.

随着 T 的增大，$f_x \to p_x^{st}$，$r_{xx'} \to k_{xx'} p_{x'}^{st}$。我们需要一个辅助的应用过程。这个过程中的 jump rate 满足 $r_{xx'} = \tilde{k}_{xx'} f_{x'}$ （对应于 $r_{xx'} = k_{xx'} \cdot p_{x'}^{st}$）.

相当于我们用经验的 $r_{xx'}$ 和 f 定下了这个过程的 $k_{xx'}$。逃逸速率 $\tilde{k}_x^{out} = \sum_{x'} \tilde{k}_{x'x}$

那么，对于每条路径来说，它在实际测度 P 下的概率与它在辅助测度 P^* 下的概率之比（换言之，两个概率测度的 R-N 导数）为：

$$\frac{P_X}{P^*_X} = \left(\prod_{i=1}^{N} \frac{k_{x_i x_{i+1}}}{\tilde{k}_{x_i x_{i+1}}}\right)\left(\prod_{i=0}^{n} \exp\left(-(k_{x_i}^{out} - \tilde{k}_{x_i}^{out})(t_{i+1}-t_i)\right)\right) \qquad (t_{n+1}=T)$$

$$= \exp\left[-T \sum_{x,x'} \left(\tilde{k}_{xx'} \ln\frac{\tilde{k}_{xx'}}{k_{xx'}} - \tilde{k}_{xx'} + k_{xx'}\right) f_x\right].$$

从而，我们有：$p(f,\tilde{k},T) = p^*(f,\tilde{k},T)\cdot \left\langle \frac{P_X}{P_X^*}\right\rangle \asymp \frac{P_X}{P_X^*} = \exp\left[-T\cdot I^{(f,\tilde{k})}(f,\tilde{k})\right]$. 其中 $I^{(f,\tilde{k})}(f,\tilde{k}) = \sum_{x,x'}\left(\tilde{k}_{xx'}\ln\frac{\tilde{k}_{xx'}}{k_{xx'}} - \tilde{k}_{xx'} + k_{xx'}\right) f_x$.

这代表用经验分布生成的 jump rate.
因此 $T\to+\infty$ 时，这一项的值趋向为 1.

利用前面的收缩原理，我们可以直接得到：$P(f,T) \asymp \exp\left[-T\cdot I^{(f)}(f)\right]$. $I^{(f)}(f) = \inf_{\tilde{k}} I^{(f,\tilde{k})}(f,\tilde{k})$.

考虑 current：$j_{xx'} = \tilde{k}_{xx'} f_x - \tilde{k}_{x'x} f_{x'}$，它对应的 rate function 可以用乘子法求出。具体而言，考虑一组反对称乘子 $\lambda_{xx'} = -\lambda_{x'x}$，我们有增广的 rate function：

$\Phi(\tilde{k}, k, \lambda) = \sum_{x,x'}\left[I(f,\tilde{k}) + \frac{1}{2}\lambda_{xx'}\left(\tilde{k}_{xx'}f_x - \tilde{k}_{x'x}f_{x'} - j_{xx'}\right)\right]$. 通过求解其极值可以得到：$I^{(f,j)}(f,j) = I^{(f,\tilde{k})}(f,\tilde{k}^*)$. $\tilde{k}^*_{xx'} = k_{xx'}\cdot\exp\left(\frac{1}{2}\lambda_{xx'}\right)$.