Part 2 Abstract theory

Probability

Entropy

Stochastic matrix and basic convergence theorem

Markov jump process

<probability>

§ basic concepts in probability theory

elementrary events

- a physical system with discrete microscopic states $j = 1, 2, \cdots, \Omega$

see part 1-p36

particle configurations, spin configurations, ...

- events $A, B, \cdots$ — A is either true or false for each state $j = 1, \cdots, \Omega$

random variables

- state quantities $\hat{f}, \hat{g}, \cdots$ — f takes value f_j in state $j = 1, \cdots, \Omega$

(may not be a standard terminology)

- probability distribution

(1) $$\mathbb{P} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_\Omega \end{pmatrix} = (p_j)_{j=1,\cdots,\Omega}$$

p_j : the probability of state j

(2) $p_j \geq 0$, (3) $\sum_{j=1}^{\Omega} p_j = 1$

- probability of an event A

(1) $$\mathrm{Prob}_{\mathbb{P}}[A] := \sum_{j=1}^{\Omega} p_j \,\chi_j[A]$$

(2) $$\chi_j[A] := \begin{cases} 1 & A \text{ is true in } j \\ 0 & A \text{ is false in } j \end{cases}$$

- expectation value of a state quantity $\hat{f}$

(3) $$\langle \hat{f} \rangle_{\mathbb{P}} := \sum_{j=1}^{\Omega} p_j f_j$$

- fluctuation (standard deviation) of $\hat{f}$

(4) $$\sigma_{\mathbb{P}}[\hat{f}] := \sqrt{\langle \hat{f}^2 \rangle_{\mathbb{P}} - (\langle \hat{f} \rangle_{\mathbb{P}})^2} = \sqrt{\left\langle \left(\hat{f} - \langle \hat{f} \rangle_{\mathbb{P}}\right)^2 \right\rangle_{\mathbb{P}}}$$

- uniform distribution

(5) $$\mathbb{P}_u = \begin{pmatrix} 1/\Omega \\ \vdots \\ 1/\Omega \end{pmatrix}$$

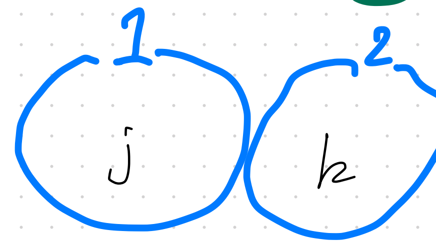
§ combined system

a system consisting of two subsystems

generalization to N subsystems is trivial

subsystem 1 : states $j=1,2,\cdots,\Omega_1$

subsystem 2 : states $k=1,2,\cdots,\Omega_2$



states of the whole system (j,k) $j=1,\cdots,\Omega_1,\ k=1,\cdots,\Omega_2$

probability distribution of the whole system $\mathbb{P}=(p_{j,k})_{\substack{j=1,\cdots,\Omega_1\\ k=1,\cdots,\Omega_2}}$

marginal distributions (1) $p^{(1)}_j := \sum_{k=1}^{\Omega_2} p_{j,k}$ (2) $p^{(2)}_k := \sum_{j=1}^{\Omega_1} p_{j,k}$

$\mathbb{P}^{(1)}, \mathbb{P}^{(2)}$ probability distributions

two subsystems are independent if $p_{j,k}=p^{(1)}_j\,p^{(2)}_k$ (3)

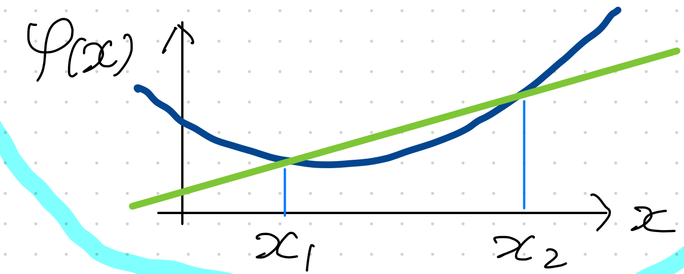
§ Jensen's inequality

$\varphi(x)$ a convex function of $x \in \mathbb{R}$

($\varphi''(x) \geq 0$ if twice differentiable)

$\forall x_1, x_2 \in \mathbb{R}$ $\quad \forall \lambda \in [0, 1]$

$$\varphi(\lambda x_1 + (1-\lambda) x_2) \leq \lambda \varphi(x_1) + (1-\lambda) \varphi(x_2) \qquad (1)$$



for any $\mathbb{P}$ and $\hat{f}$

$$(2) \quad \boxed{\varphi(\langle \hat{f} \rangle_{\mathbb{P}}) \leq \langle \varphi(\hat{f}) \rangle_{\mathbb{P}}}$$

$\varphi(\hat{f})$ takes value $\varphi(f_j)$ in state j

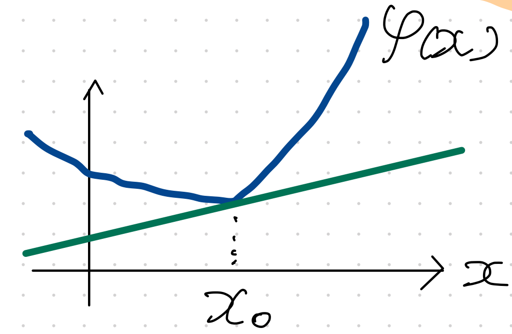
example (3) $\varphi(x) = e^{x}$

$$(4) \quad e^{\langle \hat{f} \rangle_{\mathbb{P}}} \leq \langle e^{\hat{f}} \rangle_{\mathbb{P}}$$

proof → (The same proof applies to continuous random variables)

▷ **useful fact** if $\varphi(x)$ is convex then for any x_0, there exists $\alpha \in \mathbb{R}$ s.t.

$$(1) \quad \varphi(x) \geq \varphi(x_0) + \alpha(x - x_0) \quad \text{for any } x \in \mathbb{R}$$



α : subderivative $\quad \alpha = \varphi'(x_0)$ if $\varphi(x)$ is differentiable at x_0

set $x = f_j$ and $x_0 = \langle \hat{f} \rangle_{\mathbb{P}}$ in (1)

$$(2) \quad \varphi(f_j) \geq \varphi(\langle \hat{f} \rangle_{\mathbb{P}}) + \alpha(f_j - \langle \hat{f} \rangle_{\mathbb{P}})$$

$$\sum_j p_j(\cdots) \quad \Downarrow$$

$$(3) \quad \langle \varphi(\hat{f}) \rangle_{\mathbb{P}} \geq \varphi(\langle \hat{f} \rangle_{\mathbb{P}}) + \alpha\overbrace{(\langle \hat{f} \rangle_{\mathbb{P}} - \langle \hat{f} \rangle_{\mathbb{P}})}^{=0}$$

§ canonical distribution

a probability distribution that reproduces macroscopic properties of the equilibrium state of a physical system at temperature β^{-1}

- a physical system with discrete microscopic states $j=1,2,\cdots,\Omega$
- E_j the energy of the system in the state j
- canonical distribution $\mathbb{p}^{(\mathrm{can},\beta)}$

$$\boxed{(1)\quad p_j^{(\mathrm{can},\beta)} = \frac{e^{-\beta E_j}}{Z(\beta)}}$$

partition function
$$(2)\quad Z(\beta) = \sum_{j=1}^{\Omega} e^{-\beta E_j}$$

Helmholtz free energy
$$(3)\quad F(\beta) = -\frac{1}{\beta}\log Z(\beta)$$

a system consisting of two subsystems

subsystem 1 : states $j = 1, 2, \cdots, \Omega_1$

subsystem 2 : states $k = 1, 2, \cdots, \Omega_2$

if energy of the whole system is (1) $E_{j,k} = E_j^{(1)} + E_k^{(2)}$

no interaction energy

(2)

$$Z(\beta) = \sum_{j=1}^{\Omega_1} \sum_{k=1}^{\Omega_2} e^{-\beta E_{j,k}} = \sum_{j=1}^{\Omega_1} e^{-\beta E_j^{(1)}} \sum_{k=1}^{\Omega_2} e^{-\beta E_k^{(2)}} = Z_1(\beta) Z_2(\beta)$$

(3)

$$p_{j,k}^{(can,\beta)} = \frac{e^{-\beta E_{j,k}}}{Z(\beta)} = \frac{e^{-\beta E_j^{(1)}}}{Z_1(\beta)} \frac{e^{-\beta E_k^{(2)}}}{Z_2(\beta)} = p_j^{(1,can,\beta)} \, p_k^{(2,can,\beta)}$$

two subsystems are independent

§ the physical meaning of probability

Q. What does it mean that the probability of an event A is p, e.g., $p = 0.3$?

After an experiment (trial), A may be true or false ...

Q. What does it mean that the expectation value of a physical quantity $\hat{f}$ is $\langle \hat{f} \rangle_p$?

After an experiment (trial), $\hat{f}$ takes a value f_j for some j ...

(you never get 3.5 by rolling a fair dice!)

▷ an assumption necessary for relating the probability theory with the physical world

one of many versions

Cournot's principle

Pick an event A such that $\mathrm{Prob}_p[A] \ll 1$, and make an experiment.

Then the event A is never true in practice.

Chebyshev's inequality

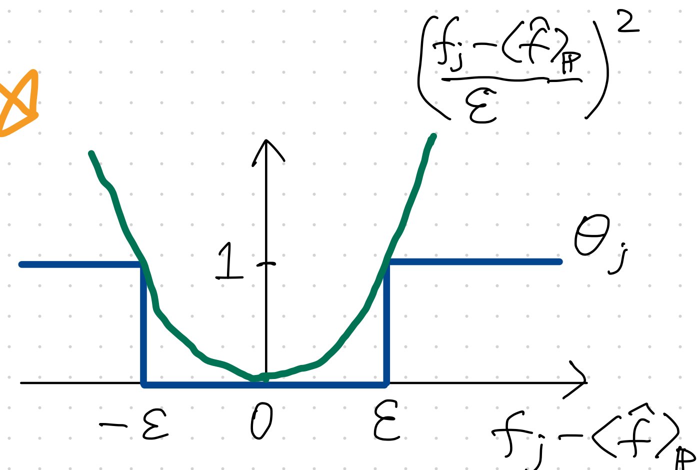
for any probability distribution $\mathbb{P}$, state quantity $\hat{f}$, and $\varepsilon > 0$

$$(1) \quad \mathrm{Prob}_{\mathbb{P}}\left[\,|\hat{f} - \langle \hat{f} \rangle_{\mathbb{P}}| \geq \varepsilon\,\right] \leq \left(\frac{\sigma_{\mathbb{P}}[\hat{f}]}{\varepsilon}\right)^2$$

proof

Define $\hat{\theta}$ by (2) $\theta_j = \begin{cases} 1 & \text{if } |f_j - \langle \hat{f} \rangle_{\mathbb{P}}| \geq \varepsilon \\ 0 & \text{if } |f_j - \langle \hat{f} \rangle_{\mathbb{P}}| < \varepsilon \end{cases}$

then (3) $\theta_j \leq \left(\dfrac{f_j - \langle \hat{f} \rangle_{\mathbb{P}}}{\varepsilon}\right)^2$ for all j



$$(4) \quad \langle \hat{\theta} \rangle_{\mathbb{P}} \leq \frac{\langle (\hat{f} - \langle \hat{f} \rangle_{\mathbb{P}})^2 \rangle_{\mathbb{P}}}{\varepsilon^2} = \left(\frac{\sigma_{\mathbb{P}}[\hat{f}]}{\varepsilon}\right)^2$$

$$\parallel$$

$$\mathrm{Prob}_{\mathbb{P}}[\,|\hat{f} - \langle \hat{f} \rangle| \geq \varepsilon\,]$$

//

Suppose (1) $\sigma_{\mathbb{P}}[\hat{f}] \ll \boxed{\varepsilon}$ → the precision for measuring $\hat{f}$

often true in macroscopic systems

Chebyshev's ineq. (2) $\mathrm{Prob}_{\mathbb{P}}\left[\,|\hat{f} - \langle\hat{f}\rangle_{\mathbb{P}}| \geq \varepsilon\,\right] \leq \left(\frac{\sigma_{\mathbb{P}}[\hat{f}]}{\varepsilon}\right)^2 \boxed{\ll 1}$

Cournot's principle ⟹ the measurement result of $\hat{f}$ is always equal to $\langle\hat{f}\rangle_{\mathbb{P}}$ within the precision!

When $\sigma_{\mathbb{P}}[\hat{f}]$ is not small

N identical independent copies of the system (3) $\hat{f}_{\mathrm{av}} := \frac{1}{N}\sum_{n=1}^{N} \boxed{\hat{f}^{(n)}}$

$\hat{f}$ for the n-th copy

(4) $\langle\hat{f}_{\mathrm{av}}\rangle_{\mathbb{P}^{\otimes N}} = \langle\hat{f}\rangle_{\mathbb{P}}$ (5) $\sigma_{\mathbb{P}^{\otimes N}}[\hat{f}_{\mathrm{av}}] = \frac{1}{\sqrt{N}}\sigma_{\mathbb{P}}[\hat{f}] \ll \varepsilon$

if N is large

The measurement result of $\hat{f}_{\mathrm{av}}$ is always equal to $\langle\hat{f}\rangle_{\mathbb{P}}$

⟨entropy⟩ → information theoretic entropy

§ Shannon entropy

system with discrete states $j=1,\cdots,\Omega$

Shannon entropy of a probability distribution $\mathbb{P}=(P_j)_{j=1,\cdots,\Omega}$

$$\boxed{S(\mathbb{P}) := -\sum_{j=1}^{\Omega} P_j \log P_j} \tag{1}$$

we use the convention $0\log 0=0$

uniform distribution $\mathbb{P}_u = \begin{pmatrix} 1/\Omega \\ \vdots \\ 1/\Omega \end{pmatrix}$

$$S(\mathbb{P}_u) = \log \Omega \tag{2}$$

in general (3) $0 \leq S(\mathbb{P}) \leq \log \Omega$

trivial

see P15-(5)

interpretation (1) $I_j = \log \frac{1}{p_j}$ information content = "amount of surprise" when the state j is observed

$\left(p_j = 1 \to \log\frac{1}{p_j} = \boxed{0} \right.$ you know that j happens, No surprise

$\left. p_j = 10^{-6} \to \log\frac{1}{p_j} = \boxed{6 \log 10} \right)$ a rare event! BIG surprise!!

$S(\mathbb{P})$ is the average of I_j

additivity system which is a combination of two subsystems

if $p_{j,k} = p_j^{(1)} p_k^{(2)}$ ← independent (2)

(3) $$S(\mathbb{P}) = -\sum_{j,k} p_{j,k} \log p_{j,k} = -\sum_j \underbrace{\boxed{\sum_k p_{j,k}}}_{= p_j^{(1)}} \log p_j^{(1)} - \sum_k \underbrace{\boxed{\sum_j p_{j,k}}}_{= p_k^{(2)}} \log p_k^{(2)}$$

$$= -\sum_j p_j^{(1)} \log p_j^{(1)} - \sum_k p_k^{(2)} \log p_k^{(2)} = S(\mathbb{P}^{(1)}) + S(\mathbb{P}^{(2)})$$

example: binary entropy

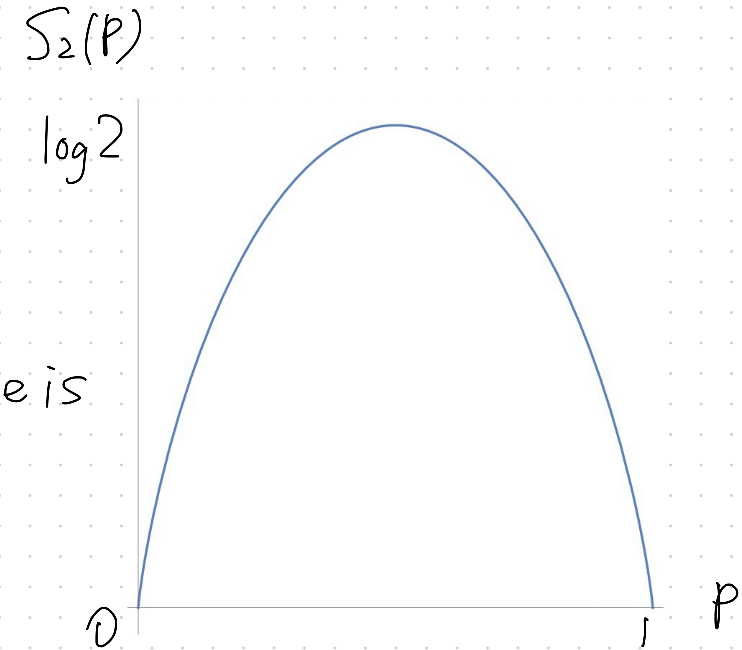
$\Omega = 2$

(1) $\mathbb{P} = \begin{pmatrix} p \\ 1-p \end{pmatrix}$

(2) $S_2(p) := S(\mathbb{P}) = -p\log p - (1-p)\log(1-p)$

(3) $S_2'(p) = \log \frac{1-p}{p}$

(4) $S_2''(p) = -\frac{1}{p(1-p)} < 0$



entropy = information = amount of surprise is

- maximum when $p = \frac{1}{2}$
- zero when $p = 0$ or 1

§ relative entropy a.k.a. KL divergence

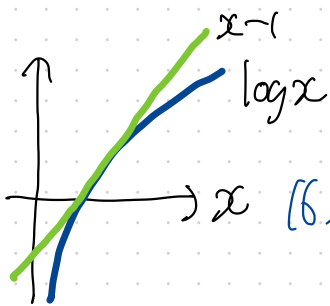
$\mathbb{P}, \mathbb{Q}$ probability distributions

relative entropy or Kullback-Leibler divergence

(1) $\boxed{D(\mathbb{P}\|\mathbb{Q}) := \sum_{j=1}^{\Omega} p_j \log \frac{p_j}{q_j}}$ ($D(\mathbb{P}\|\mathbb{Q}) = \infty$ if $p_j \neq 0$ and $q_j = 0$ for some j)

basic property (2) $D(\mathbb{P}\|\mathbb{Q}) \geq 0$ (3) $D(\mathbb{P}\|\mathbb{Q}) = 0 \iff \mathbb{P} = \mathbb{Q}$

proof recall that (4) $\log x \leq x - 1$ for $x > 0$



thus (5) $\log \frac{1}{x} \geq 1 - x$ ($\log \frac{1}{x} > 1 - x$ if $x \neq 1$)

(6) $D(\mathbb{P}\|\mathbb{Q}) \geq \sum_j p_j \left(1 - \frac{q_j}{p_j}\right) = \sum_j p_j - \sum_j q_j = 0$

$>$ for at least one j if $\mathbb{P} \neq \mathbb{Q}$ //

(1) $D(\mathbb{P}|\mathbb{Q}) := \sum_{j=1}^{\Omega} p_j \log \frac{p_j}{q_j}$

(2) $D(\mathbb{P}|\mathbb{Q}) \geq 0$ (3) $D(\mathbb{P}|\mathbb{Q}) = 0 \iff \mathbb{P} = \mathbb{Q}$

$D(\mathbb{P}|\mathbb{Q})$ is an asymmetric distance between $\mathbb{P}$ and $\mathbb{Q}$

uniform distribution $\mathbb{P}_u = \begin{pmatrix} 1/\Omega \\ \vdots \\ 1/\Omega \end{pmatrix}$ (4) $D(\mathbb{P}|\mathbb{P}_u) = \sum_j p_j (\log p_j + \log \Omega)$

$= \log \Omega - S(\mathbb{P})$

Shannon entropy is essentially the KL-divergence with respect to $\mathbb{P}_u$

(5) $D(\mathbb{P}|\mathbb{P}_u) \geq 0 \implies \log \Omega \geq S(\mathbb{P})$ p11-(3)

when $\mathbb{P}$ and $\mathbb{Q}$ are close

(6) $D(\mathbb{P}|\mathbb{Q}) = \frac{1}{2} \sum_{j=1}^{\Omega} \frac{(p_j - q_j)^2}{q_j} + O(|\mathbb{P}-\mathbb{Q}|^3)$

§ relations to statistical mechanics

Shannon entropy and statistical mechanical entropy

fix E_j $(j=1,\ldots,\Omega)$

canonical distribution (1) $$p_j^{(\mathrm{can},\beta)} = \frac{e^{-\beta E_j}}{Z(\beta)}$$

$$\boxed{\langle \hat{f} \rangle_\beta^{\mathrm{can}} = \langle \hat{f} \rangle_{\mathbb{P}^{(\mathrm{can},\beta)}}}$$

(2) $$S(\mathbb{P}^{(\mathrm{can},\beta)}) = -\sum_{j=1}^{\Omega} p_j^{(\mathrm{can},\beta)} \log \frac{e^{-\beta E_j}}{Z(\beta)}$$

$$= \sum_{j=1}^{\Omega} p_j^{(\mathrm{can},\beta)} \{\beta E_j + \log Z(\beta)\}$$

$$= \beta \langle \hat{E} \rangle_\beta^{\mathrm{can}} + \log Z(\beta) = \beta\{\langle \hat{E} \rangle_\beta^{\mathrm{can}} - F(\beta)\}$$

$$= \frac{1}{T}\{\langle \hat{E} \rangle_\beta^{\mathrm{can}} - F(\beta)\} = \boxed{S(\beta)}$$ ← statistical mechanical entropy

Variational characterization of the canonical distribution

for any probability distribution $\mathbb{P}$

(1) $$D(\mathbb{P}\,|\,\mathbb{P}^{(\mathrm{can},\beta)}) = \sum_{j=1}^{\Omega} p_j \log\left(p_j \frac{Z(\beta)}{e^{-\beta E_j}}\right)$$

$$= \sum_j p_j \log p_j + \beta \sum_j p_j E_j + \underbrace{\log Z(\beta)}_{= -\beta F(\beta)}$$

$$= \underbrace{-S(\mathbb{P}) + \beta \langle \hat{E} \rangle_{\mathbb{P}}}_{\text{define } \beta F(\mathbb{P})} - \beta F(\beta) \geq 0$$

$=$ only when $\mathbb{P} = \mathbb{P}^{(\mathrm{can},\beta)}$

(2) $$F(\mathbb{P}) \geq F(\beta)$$

$\mathbb{P}^{(\mathrm{can},\beta)}$ is the unique probability distribution that minimizes

(3) $$F(\mathbb{P}) = \langle \hat{E} \rangle_{\mathbb{P}} - \frac{1}{\beta} S(\mathbb{P})$$

"Helmholtz free energy" for general $\mathbb{P}$

<stochastic matrix and basic convergence theorem>

§ stochastic matrix

$\Omega \times \Omega$ matrix (1) $T = (T_{j,k})_{j,k=1,\cdots,\Omega}$

such that (2) $T_{j,k} \geq 0$ and (3) $\displaystyle\sum_{j=1}^{\Omega} T_{j,k} = 1$ for all k

if $\mathbb{P}$ is a probability distribution

then (4) $\mathbb{P}' = T\mathbb{P}$ is also a probability distribution

proof

$$(5)\quad p_j' = \sum_{k=1}^{\Omega} T_{jk}\, p_k \Rightarrow (6)\quad p_j' \geq 0$$

$$\Downarrow$$

$$(7)\quad \sum_{j=1}^{\Omega} p_j' = \sum_{k=1}^{\Omega} \overbrace{\boxed{\sum_{j=1}^{\Omega} T_{j,k}}}^{1} p_k = \sum_{k=1}^{\Omega} p_k = 1$$

§ monotonicity of the KL-divergence

$\mathbb{P}, \mathbb{Q}$ arbitrary probability distributions, T arbitrary stochastic matrix

$$\boxed{(1)\quad D(\mathbb{P}\,|\,\mathbb{Q}) \geq D(T\mathbb{P}\,|\,T\mathbb{Q})}$$

proof (2) $p_j' = \sum_k T_{jk}\, p_k$, $\quad q_j' = \sum_k T_{jk}\, q_k$

define (3) $\tilde{p}_j^{(k)} = \dfrac{T_{kj}\, p_j}{p_k'}$, $\quad \tilde{q}_j^{(k)} = \dfrac{T_{kj}\, q_j}{q_k'}$

then (4) $\sum_j \tilde{p}_j^{(k)} = \dfrac{p_k'}{p_k'} = 1$

$\tilde{\mathbb{P}}^{(k)}, \tilde{\mathbb{Q}}^{(k)}$ are probability distributions

(4)

$$D(\mathbb{P}\,|\,\mathbb{Q}) - D(\mathbb{P}'\,|\,\mathbb{Q}') = \sum_j p_j \log\frac{p_j}{q_j} - \sum_j p_j' \log\frac{p_j'}{q_j'}$$

$$= \sum_{k,j} T_{kj}\, p_j \log\frac{T_{kj}\, p_j}{T_{kj}\, q_j} - \sum_j p_j' \log\frac{p_j'}{q_j'}$$

$$\left\{\begin{aligned} T_{kj}\, p_j &= \tilde{p}_j^{(k)}\, p_k' \\ T_{kj}\, q_j &= \tilde{q}_j^{(k)}\, q_k' \end{aligned}\right.$$

$$= \sum_{k,j} p_k'\, \tilde{p}_j^{(k)} \log\frac{\tilde{p}_j^{(k)}}{\tilde{q}_j^{(k)}} + \underbrace{\sum_{k,j} p_k'\, \tilde{p}_j^{(k)} \log\frac{p_k'}{q_k'} - \sum_j p_j' \log\frac{p_j'}{q_j'}}_{=0}$$

$$= \sum_k p_k'\, D(\tilde{\mathbb{P}}^{(k)}\,|\,\tilde{\mathbb{Q}}^{(k)}) \geq 0 \qquad //$$

§ Convergence theorem

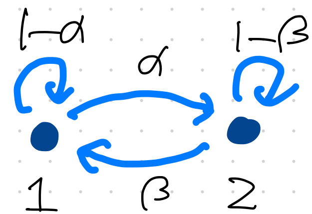
T stochastic matrix, $\mathbb{P}^{(0)}$ arbitrary probability distribution

how does the prob. dist. $T^n \mathbb{P}^{(0)}$ behave for large n?

(Markov process with discrete time $n=1,2,\cdots$)

example with $\Omega = 2$

$$\boxed{\begin{array}{ll} T_{11} = 1-\alpha & T_{12} = \beta \\ T_{21} = \alpha & T_{22} = 1-\beta \end{array}}$$



(1) $$T = \begin{pmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{pmatrix} \qquad 0<\alpha<1, \quad 0<\beta<1$$

(2) $$\mathbb{P}^{(0)} = \begin{pmatrix} p \\ 1-p \end{pmatrix} \quad (0 \leq p \leq 1)$$

(3) $$T\mathbb{P}^{(0)} = \begin{pmatrix} (1-\alpha-\beta)p + \beta \\ (\alpha+\beta-1)p + (1-\beta) \end{pmatrix}, \quad \cdots$$

one can compute $T^n \mathbb{P}^{(0)}$ for general n by diagonalizing T

(4) $$\det(T-\lambda I) = \det\begin{pmatrix} 1-\alpha-\lambda & \beta \\ \alpha & 1-\beta-\lambda \end{pmatrix} = \lambda^2 + (\alpha+\beta-2)\lambda + (1-(\alpha+\beta))$$

$$= (\lambda - 1)(\lambda - \{1-(\alpha+\beta)\})$$

(1) $T = \begin{pmatrix} 1-\alpha & \beta \\ \alpha & 1-\beta \end{pmatrix}$ $\begin{cases} \text{eigenvalues (2) } \lambda_1 = 1,\ \lambda_2 = 1-(\alpha+\beta) \\ \text{eigenvectors (3) } v_1 = \begin{pmatrix} \frac{\beta}{\alpha+\beta} \\ \frac{\alpha}{\alpha+\beta} \end{pmatrix},\ v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \end{cases}$

normalized as a probability distribution

then (4) $\mathbb{P}^{(0)} = \begin{pmatrix} p \\ 1-p \end{pmatrix} = v_1 + \left(p - \frac{\beta}{\alpha+\beta}\right) v_2$

and (5) $T^n \mathbb{P}^{(0)} = T^n v_1 + \left(p - \frac{\beta}{\alpha+\beta}\right) T^n v_2$

$$= v_1 + \left(p - \frac{\beta}{\alpha+\beta}\right) \{1-(\alpha+\beta)\}^n v_2$$

since $|1-(\alpha+\beta)| < 1$

(6) $\lim_{n\uparrow\infty} T^n \mathbb{P}^{(0)} = \boxed{v_1}$ for any probability distribution $\mathbb{P}^{(0)} = \begin{pmatrix} p \\ 1-p \end{pmatrix}$

converges to a unique probability distribution

Convergence theorem

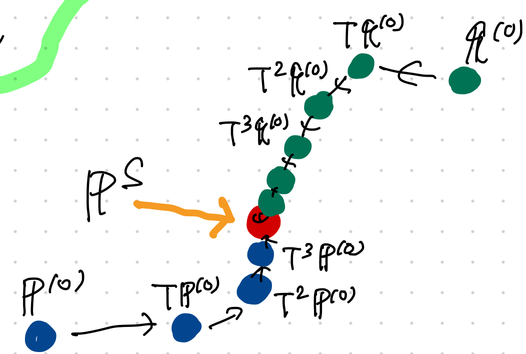
Definition: a stochastic matrix T is said to be primitive (or irreducible + aperiodic) if there is an integer $\nu \geq 1$ such that (1) $(T^{\nu})_{j,k} > 0$ for any j, k.

Theorem: assume that T is primitive. then

- there is a unique probability distribution $\mathbb{P}^S = (P^S_j)_{j=1,\dots,\Omega}$ that satisfies (2) $T\,\mathbb{P}^S = \mathbb{P}^S$
- it holds that $P^S_j > 0$ for any j
- for any prob. distribution $\mathbb{P}^{(0)}$, we have

(3) $\boxed{\lim_{n \to \infty} T^n \mathbb{P}^{(0)} = \mathbb{P}^S}$

stationary distribution



Convergence to a unique stationary distribution is a univeral phenomenon!

▷ proof

(1) $V_0 := \{ \psi = (\psi_j)_{j=1,\cdots,\Omega} \in \mathbb{R}^{\Omega} \mid \sum_{j=1}^{\Omega} \psi_j = 0 \}$ → (ψ_2 in p21 belongs to V_0)

Lemma: if T is primitive then (2) $\lim_{n\to\infty} T^n \psi = 0$ for any $\psi \in V_0$

proof of Theorem given Lemma

(3) $\left(T^t \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}\right)_j = \sum_{k=1}^{\Omega} (T^t)_{jk} = \sum_{k=1}^{\Omega} T_{kj} = 1 \Rightarrow$ (4) $T^t \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$ $\lambda = 1$ is an eigenvalue of T^t

$\lambda = 1$ is an eigenvalue of T → corresponding eigenvector $u = (u_j)_{j=1,\cdots,\Omega}$ with $u_j \in \mathbb{R}$

(5) $Tu = u \xrightarrow{\text{Lemma}} u \notin V_0 \Leftrightarrow$ (6) $\sum_{j=1}^{\Omega} u_j \neq 0$

define $\mathbb{P}^S$ by (7) $p_j^S = \left(\sum_{k=1}^{\Omega} u_k\right)^{-1} u_j$ then (8) $\sum_{j=1}^{\Omega} p_j^S = 1$

we still do not know whether $\mathbb{P}^S$ is a probability distribution or not there is $\mathbb{P}^S = (P_j^S)_{j=1,\cdots,\Omega}$ such that (1) $T\,\mathbb{P}^S = \mathbb{P}^S$ (2) $\sum_{j=1}^{\Omega} P_j^S = 1$

- for arbitrary probability distribution $\mathbb{P}^{(0)}$

$$(3)\quad T^n \mathbb{P}^{(0)} = \underbrace{T^n \mathbb{P}^S}_{=\mathbb{P}^S} + T^n \underbrace{(\mathbb{P}^{(0)} - \mathbb{P}^S)}_{\in V_0} \xrightarrow[\text{Lemma}]{n\nearrow\infty} \mathbb{P}^S$$

convergence is proved.

- since $T^n \mathbb{P}^{(0)}$ is a probability distribution, so is $\mathbb{P}^S$

$$(4)\quad \mathbb{P}^S = T^\nu \mathbb{P}^S \rightarrow (5)\quad P_j^S = \sum_{k=1}^{\Omega} \underbrace{(T^\nu)_{jk}}_{\text{positive}} \underbrace{P_k^S}_{\text{nonnegative}} > 0$$

properties of $\mathbb{P}^S$ are proved.

- assume $T\,\mathbb{P}^A = \mathbb{P}^A$ for a prob. dist. $\mathbb{P}^A \neq \mathbb{P}^S$

  this contradicts with (3) if we set $\mathbb{P}^{(0)} = \mathbb{P}^A$

uniqueness of the solution of $T\mathbb{P}^S = \mathbb{P}^S$ is proved. //

(remark $\lambda = 1$ and $\mathbb{P}^{(0)}$ are the Perron-Frobenius eigenvalue and eigenvector of T. We avoided the use of the Perron-Frobenius theorem)

Perron–Frobenius theorem

let $A = (a_{jk})_{j,k=1,\dots,\Omega}$ be an $\Omega \times \Omega$ matrix with

(i) $a_{jk} \in \mathbb{R}$ for any j, k

(ii) $a_{jk} \geq 0$ for any $j \neq k$

j and k are "connected" via nonzero entries

(iii) A is irreducible, i.e., for any $j \neq k$, there exist $n > 0$ and $i_0, i_1, \dots, i_n$ s.t. $i_0 = j$, $i_n = k$, and $a_{i_{\ell-1} i_\ell} \neq 0$ for $\ell = 1, \dots, n$

then there exists a nondegenerate real eigenvalue λ^{PF} of A, and the corresponding eigenvector $v^{PF} = (v_j^{PF})_{j=1,\dots,\Omega}$ can be chosen to satisfy $v_j^{PF} > 0$ for all j.

any other eigenvalue λ of A satisfies $\mathrm{Re}\,\lambda < \lambda^{PF}$

proof of Lemma

- some definitions

(1) $\mu = \min_{j,k} (T^\nu)_{j,k} > 0$

M : $\Omega \times \Omega$ matrix such that $(M)_{j,k} = \mu$ for all j,k

(2) $S = T^\nu - M$ $\quad (0 \leq 1 - \Omega\mu < 1)$

then (3) $(S)_{j,k} = (T^\nu)_{j,k} - \mu \geq 0$ (4) $\sum_{j=1}^{\Omega} (S)_{j,k} = 1 - \Omega\mu$

- for any $v \in V_0$

(5) $\sum_{j=1}^{\Omega} (T^\nu v)_j = \sum_{j,k} (T^\nu)_{jk} v_k = \sum_k v_k = 0 \rightarrow$ (6) $T^\nu v \in V_0$

(7) $(M v)_j = \sum_k (M)_{j,k} v_k = \mu \sum_k v_k = 0$

we thus have (8) $M v = 0$ and hence (9) $T^\nu v = S v$

L^1 norm of a vector $w = (w_j)_{j=1,\cdots,\Omega} \in \mathbb{R}^{\Omega}$ (1) $\|w\|_1 := \sum_{j=1}^{\Omega} |w_j|$

for any $v \in V_0$

$$(2)\quad \|T^\nu v\|_1 = \sum_j |(T^\nu v)_j| \overset{\text{p26-(9)}}{=} \sum_j |(Sv)_j| \leq \sum_{j,k} S_{jk} |v_k|$$

$$\overset{\text{p26-(4)}}{=} (1-\Omega\mu) \sum_k |v_k| = (1-\Omega\mu)\|v\|_1$$

since $T^\nu v \in V_0$, we can repeatedly use (2) to get

$$(3)\quad \|T^{\nu m} v\|_1 \leq (1-\Omega\mu)^m \|v\|_1$$

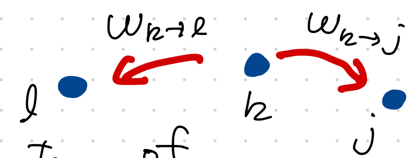
thus

$$(4)\quad \lim_{m\uparrow\infty} \|T^{\nu m} v\|_1 = 0 \Rightarrow (5)\quad \lim_{m\uparrow\infty} T^{\nu m} v = \mathbb{O}$$

<Markov jump process>

a Markov process with discrete states and continous time $t \geq 0$

§ definitions



- microscopic states $j = 1, 2, \cdots, \Omega$
- a Markov jump process is fully characterized by a collection of transition rate $w_{k\to j}(t) \geq 0$ for $k, j = 1, \cdots, \Omega$, $k \neq j$, and $t \geq 0$
- the collection of transition rates at t $\quad w(t) = (w_{k\to j}(t))_{\substack{k,j=1,\cdots,\Omega \\ (k\neq j)}}$

  the collection of $w(t)$ over whole $t \geq 0$ $\quad \tilde{w} = (w(t))_{t\geq 0}$
- if the system is in state k at time t, then it is in state j at time $t + \Delta t$ with probability $\Delta t\, w_{k\to j}(t) + O((\Delta t)^2)$ $\quad (\Delta t > 0)$
- escape rate (1) $\lambda_k(t) = \sum_{j(\neq k)} w_{k\to j}(t) \geq 0$

  if the system is in state k at time t, then the probability that it is no longer in k at time $t + \Delta t$ is $\Delta t\, \lambda_k(t) + O((\Delta t)^2)$

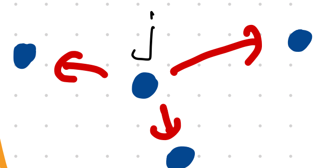
§ master equation

$\mathbb{P}(t) = (P_j(t))_{j=1,\cdots,\Omega}$ probability distribution at time t

$P_j(t)$ the probability that the system is in state j at time t

(1)
$$P_j(t+\Delta t) - P_j(t) = \boxed{-\{\Delta t\, \lambda_j(t) + O((\Delta t)^2)\}\, P_j(t)} + \sum_{k(\neq j)} \boxed{\{\Delta t\, \omega_{k\to j}(t) + O((\Delta t)^2)\}\, P_k(t)}$$

escape from j



jump from k to j

by letting $\Delta t \downarrow 0$

(2)
$$\dot{P}_j(t) = \boxed{-\lambda_j(t)\, P_j(t)} + \sum_{k(\neq j)} \boxed{\omega_{k\to j}(t)\, P_k(t)}$$

(1) $\dot{P}_j(t) = \boxed{-\lambda_j(t) P_j(t)} + \sum_{k(\neq j)}{}' \boxed{w_{k\to j}(t)\, P_k(t)}$

define transition rate matrix by (2) $R(t) = (R_{jk}(t))_{j,k=1,\cdots,\Omega}$

$$\begin{cases} (3)\ R_{jk}(t) = w_{k\to j}(t) \geq 0 \quad (j \neq k) \\ (4)\ R_{kk}(t) = -\lambda_k(t) = -\sum_{j(\neq k)}{}' w_{k\to j}(t) \leq 0 \end{cases}$$

any $R(t)$ with $R_{jk}(t) \geq 0$ for $j\neq k$ and (5) is a transition rate matrix

we then have (5) $\sum_{j=1}^{\Omega}{}' R_{jk}(t) = 0$ for any k

(1) is written as

(6) $\dot{P}_j(t) = \sum_{k=1}^{\Omega}{}' R_{jk}(t) P_k(t)$

or

(7) $\dot{\mathbb{P}}(t) = R(t)\, \mathbb{P}(t)$

master equation or Kolmogorov's forward equation

determines $\mathbb{P}(t)$ for $t \geq 0$, given $\mathbb{P}(0)$

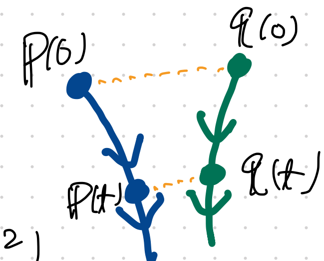
monotonicity

suppose $\mathbb{P}(t)$ and $\mathbb{Q}(t)$ satisfy

(1) $\dot{\mathbb{P}}(t) = R(t)\,\mathbb{P}(t) \qquad \dot{\mathbb{Q}}(t) = R(t)\,\mathbb{Q}(t)$

with common transition rate matrix $R(t)$

then $D(\mathbb{P}(t)|\mathbb{Q}(t))$ is non-increasing in t



proof for $t>0$ and small $\varepsilon>0$

(2) $\mathbb{P}(t+\varepsilon) = \mathbb{P}(t) + \varepsilon R(t)\,\mathbb{P}(t) + O(\varepsilon^2) = T\,\mathbb{P}(t) + O(\varepsilon^2)$

where (3) $T = I + \varepsilon R(t)$ is a stochastic matrix

(4) $\mathbb{Q}(t+\varepsilon) = T\,\mathbb{Q}(t) + O(\varepsilon^2)$

then

p19-(1)

(5) $D(\mathbb{P}(t+\varepsilon)|\mathbb{Q}(t+\varepsilon)) - D(\mathbb{P}(t)|\mathbb{Q}(t)) = D(T\mathbb{P}(t)|T\mathbb{Q}(t)) - D(\mathbb{P}(t)|\mathbb{Q}(t)) + O(\varepsilon^2) \leq O(\varepsilon^2)$

(6) $\frac{d}{dt} D(\mathbb{P}(t)|\mathbb{Q}(t)) \leq 0$

probability current $j_{j\to k}(t)$ for $j\neq k$

(1) $$j_{j\to k}(t) := \underbrace{R_{kj}(t)\,P_j(t)}_{j\text{ to }k} - \underbrace{R_{jk}(t)\,P_k(t)}_{k\text{ to }j}$$

the net flow of probability from j to k.

(2) $$j_{j\to k}(t) = -j_{k\to j}(t)$$

(3) $$\sum_{k(\neq j)} j_{j\to k}(t) = \overbrace{\left\{\sum_{k(\neq j)} R_{kj}(t)\right\}}^{=\,-R_{jj}(t)} P_j(t) - \sum_{k(\neq j)} R_{jk}(t)\,P_k(t)$$

$$= -\sum_{k=1}^{\Omega} R_{jk}(t)\,P_k(t) = -\dot{P}_j(t)$$

master equation

we thus have the continuity equation

(4) $$\dot{P}_j(t) + \sum_{k(\neq j)} j_{j\to k}(t) = 0$$

§ Convergence theorem for stationary proces ($w_{j\to k}(t)$ is independent of t)

time-independent transition rate matrix $R = (R_{jk})_{j,k=1,\dots,\Omega}$

master equation (1) $\dot{\mathbb{P}}(t) = R\,\mathbb{P}(t)$

$\rightarrow$ solution (2) $\mathbb{P}(t) = e^{tR}\,\mathbb{P}(0)$

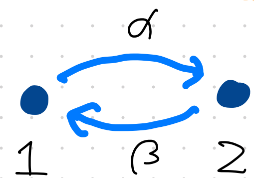
$$(3)\quad e^{A} = \sum_{n=0}^{\infty} \frac{1}{n!} A^{n} = \lim_{N\nearrow\infty} \left(I + \frac{A}{N}\right)^{N}$$

$$(4)\quad \frac{d}{dt} e^{tR} = R\,e^{tR}$$

example with $\Omega = 2$

$$R = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix} \qquad \begin{pmatrix} R_{11} = -\alpha & R_{12} = \beta \\ R_{21} = \alpha & R_{22} = -\beta \end{pmatrix}$$

$\alpha > 0,\ \beta > 0$



$$\begin{cases} (5)\quad \dot{P}_1(t) = -\alpha P_1(t) + \beta P_2(t) \\ (6)\quad \dot{P}_2(t) = \alpha P_1(t) - \beta P_2(t) \end{cases}$$

solution

$$\begin{cases} (7)\quad P_1(t) = C e^{-(\alpha+\beta)t} + \dfrac{\beta}{\alpha+\beta} \\ (8)\quad P_2(t) = -C e^{-(\alpha+\beta)t} + \dfrac{\alpha}{\alpha+\beta} \end{cases}$$

Thus

convergence to a stationary distribution!

$$(9)\quad \lim_{t\uparrow\infty} \mathbb{P}(t) = \begin{pmatrix} \dfrac{\beta}{\alpha+\beta} \\ \dfrac{\alpha}{\alpha+\beta} \end{pmatrix}.$$

for $\forall\, \mathbb{P}(0)$

Convergence theorem

Definition: a transition rate matrix R is said to be irreducible if, for any $j \neq k$ there exist $n > 0$ and $i_0, i_1, \dots, i_n$ s.t. $i_0 = j$, $i_n = k$, and $R_{i_l i_{l-1}} > 0$ for all $l = 1, \dots, n$

any j and k are "connected" via nonzero entries

→ p.25

Theorem: assume that R is irreducible then

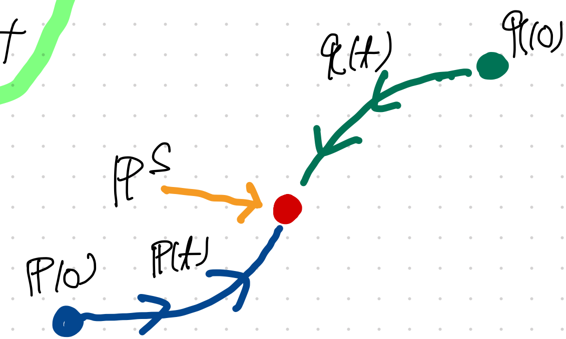
- there is a unique probability distribution $\mathbb{P}^S = (P_j^S)_{j=1,\dots,\Omega}$ that satisfies (1) $R\mathbb{P}^S = 0$

  stationary distribution

- it holds that $P_j^S > 0$ for any j
- for any initial distribution $\mathbb{P}(0)$ it holds that

$$(2) \quad \boxed{\lim_{t \uparrow \infty} \mathbb{P}(t) = \mathbb{P}^S}$$

$\left(\mathbb{P}(t) \text{ solution of } \dot{\mathbb{P}}(t) = R\,\mathbb{P}(t) \right)$



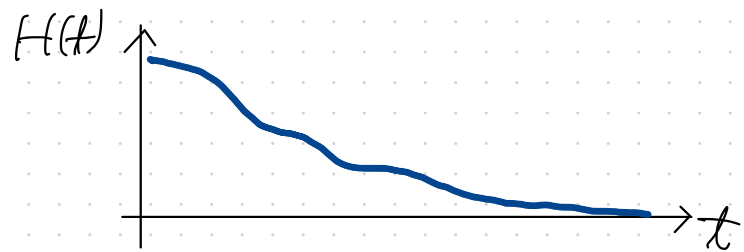
general "H-theorem" for Markov jump processes

historical name, given by Boltzmann

Corollary: suppose that R is irreducible
define (2) $H(\mathbb{P}) := D(\mathbb{P} \| \mathbb{P}^s)$
then, for any $\mathbb{P}(0)$, $H(\mathbb{P}(t))$ is non-increasing in $t \geq 0$ and
converges to zero as $t \uparrow \infty$

($\mathbb{P}(t)$ solution of $\dot{\mathbb{P}}(t) = R\,\mathbb{P}(t)$)

there is a function that knows the "arrow of time"



Proof of the theorem

Lemma: for any $\tau>0$, $T=e^{\tau R}$ is a stochastic matrix, and satisfies: $(T)_{kj}>0$ for any j,k

more or less trivial ...

Proof of Theorem given Lemma

fix $\tau>0$ (we need Lemma only for a single value $\tau>0$)

$T=e^{\tau R}$ is primitive $\rightarrow$ convergence theorem for $T^n \mathbb{P}(0)$ $\rightarrow$ P22

unique $\mathbb{P}^S$ s.t. (1) $T\mathbb{P}^S=\mathbb{P}^S$ $\Leftrightarrow$ (2) $R\mathbb{P}^S=0$

write $t=m\tau+s$ with $0\leq s<\tau$

$$\mathbb{P}(t)=e^{(m\tau+s)R}\,\mathbb{P}(0)=e^{sR}\,T^m\,\mathbb{P}(0)$$

converges to $\mathbb{P}^S$ as $m\uparrow\infty$

//

proof of Lemma

▷ proof of (1) $\sum_{k=1}^{\Omega}(e^{\tau R})_{kj}=1$

(2) $\sum_{k=1}^{\Omega}(e^{0R})_{kj}=1$

(3) $\frac{d}{d\tau}\sum_{k=1}^{\Omega}(e^{\tau R})_{kj}=\sum_{k,l=1}^{\Omega}R_{kl}(e^{\tau R})_{lj}=0$

▷ proof of $(e^{\tau R})_{kj}>0$ for any $\tau>0$ and $j,k=1,\cdots,\Omega$

- diagonal

(4) $(e^{sR})_{jj}=1+\boxed{\sum_{n=1}^{\infty}\frac{(R^n)_{jj}}{n!}s^n}=f(s)$

$f(s)$ is continuous in s and $f(0)=0$

(5) $(e^{sR})_{jj}>0$ for sufficiently small $s\geq 0$

Definition: a transition rate matrix R is said to be irreducible if, for any $j \neq k$ there exist $n > 0$ and $i_0, i_1, \dots, i_n$ s.t. $i_0 = j$, $i_n = k$, and $R_{i_\ell i_{\ell-1}} > 0$ for all $\ell = 1, \dots, n$

minimum n with the above property

38

- off-diagonal for any $j \neq k$ there is $n_0 = 1, 2, \cdots$ such that

(2) $(R^{n_0})_{kj} > 0, \quad (R^n)_{kj} = 0$ for $n < n_0$

$\to 0$ as $s \downarrow 0$

(3) $$(e^{sR})_{kj} = \sum_{n=0}^{\infty} \frac{(R^n)_{kj}}{n!} s^n = \left\{ \frac{(R^{n_0})_{kj}}{n_0!} + \boxed{\sum_{n>n_0} \frac{(R^n)_{kj}}{n!} s^{n-n_0}} \right\} s^{n_0}$$

> 0 for sufficiently small $s > 0$

$$\boxed{\exists \tau_0 > 0 \text{ s.t. } (e^{sR})_{kj} > 0 \text{ for any } j, k \text{ if } 0 < s \leq \tau_0}$$

(4) $e^{\tau R} = \left(e^{\frac{\tau}{N} R}\right)^N \Rightarrow \boxed{(e^{\tau R})_{kj} > 0 \text{ for any } j, k \text{ and any } \tau > 0}$

§ description in terms of path

- arbitrary Markov jump process with time-dependent transition rates $W(t) = (w_{j\to k}(t))_{j,k=1,\ldots,\Omega}$ $(j \neq k)$
- escape rate $\lambda_j(t) = \sum_{k(\neq j)}' w_{j\to k}(t)$
- the collection of transition rates $\tilde{w} = (W(t))_{t\in[0,\tau]}$

($\tau > 0$ final time)

staying probability $\tilde{p}_j(t,t')$ $(t \leq t')$

the probability that the system is always in state j in the time interval $[t,t']$, provided that it is in j at time t

then (1) $\tilde{p}_j(t,t) = 1$ (2) $\frac{\partial}{\partial t'}\tilde{p}_j(t,t') = -\lambda_j(t')\tilde{p}_j(t,t')$

(3) $\tilde{p}_j(t,t') = \exp\left[-\int_t^{t'} ds\, \lambda_j(s)\right]$

40

▷ path γ: a path (or a history) in the time interval $[0, \tau]$ that connects states γ_{init} and γ_{fin} (with n jumps)

(1) $\gamma = (\underbrace{j_0}_{\gamma_{init}}, j_1, \ldots, \underbrace{j_n}_{\gamma_{fin}}; t_1, \ldots, t_n)$ with (2) $\overset{t_0}{0} < t_1 < t_2 < \cdots < t_n < \overset{t_{n+1}}{\tau}$

$n = 0, 1, 2, \cdots$

m-th jump at t_m from state j_{m-1} to j_m

we write (3) $\gamma(t) = j_m$ for $t \in (t_m, t_{m+1}]$ and $\gamma(0) = j_0$

▷ transition probability density of a path γ

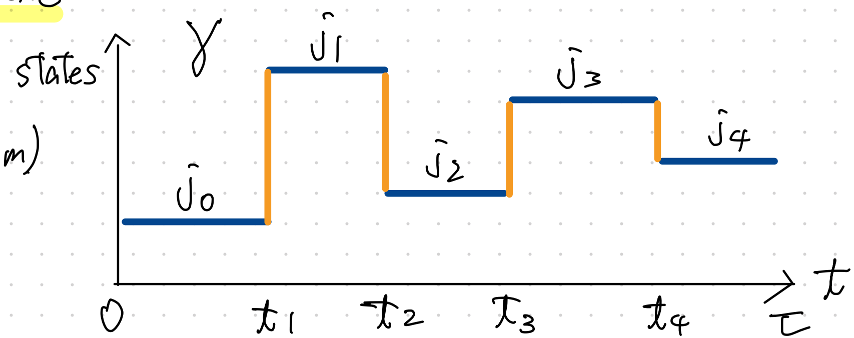
(4) $\mathcal{J}_{\tilde{\omega}}(\gamma) := \tilde{p}_{j_0}(0, t_1)\, \omega_{j_0 \to j_1}(t_1)\, \tilde{p}_{j_1}(t_1, t_2)\, \omega_{j_1 \to j_2}(t_2) \cdots \omega_{j_{n-1} \to j_n}(t_n)\, \tilde{p}_{j_n}(t_n, \tau)$

staying probability

transition rate

$$= \prod_{m=0}^{n} \tilde{p}_{j_m}(t_m, t_{m+1}) \prod_{m=1}^{n} \omega_{j_{m-1} \to j_m}(t_m)$$

$t_0 = 0, \quad t_{n+1} = \tau$



(1) $\mathbb{P}(0) = (P_j(0))_{j=1,\cdots,\Omega}$ arbitrary initial probability distribution

▷ probability density that a path γ is realized

(2) $P_{\gamma_{init}}(0)\, \mathcal{J}_{\tilde{\omega}}(\gamma) = \underbrace{P_{j_0}(0)}_{\text{initial distribution}}\, \underbrace{\tilde{P}_{j_0}(0,t_1)}_{\text{staying}}\, \underbrace{\omega_{j_0\to j_1}(t_1)}_{\text{jump}} \cdots \underbrace{\omega_{j_{n-1}\to j_n}(t_n)}_{\text{jump}}\, \underbrace{\tilde{P}_{j_n}(t_n,\tau)}_{\text{staying}}$

$\gamma_{init} = j_0$

normalization (3) $\int D\gamma\, P_{\gamma_{init}}(0)\, \mathcal{J}_{\tilde{\omega}}(\gamma) = 1$

where the "sum" over all possible paths is

(4) $\int D\gamma (\cdots) = \sum_{n=0}^{\infty} \boxed{\sum_{j_0,j_1,\cdots,j_n=1}^{\Omega}{}'} \int_0^\tau dt_1 \int_{t_1}^\tau dt_2 \int_{t_2}^\tau dt_3 \cdots \int_{t_{n-1}}^\tau dt_n (\cdots)$

$\sum'$ indicates the constraint $j_{m-1} \neq j_m$ $(m=1,\cdots,n)$

it holds that

(5) $P_j(t) = \int D\gamma\, P_{\gamma_{init}}(0)\, \mathcal{J}_{\tilde{\omega}}(\gamma)\, \delta_{\gamma(t),j}$ $\quad t\in[0,\tau]$

Remark: Key observation for the relations in p42

γ any path for the interval $[0,\tau]$ $\qquad \gamma(\tau)=j$

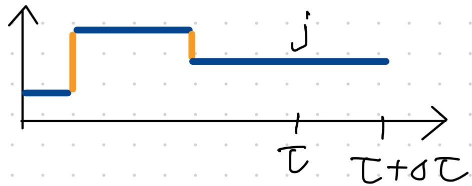
γ' path on $[0,\tau+\Delta\tau]$ such that $\gamma'(t)=\gamma(t)$ for $t\in[0,\tau]$

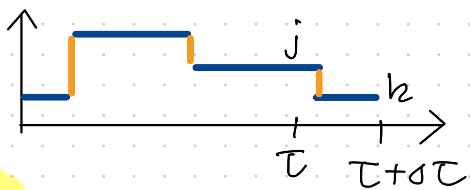
(probability density for γ') $= P_{\gamma_{\mathrm{init}}}(0)\, J_{\tilde{w}}(\gamma')$ $\qquad \left(\tilde{p}_j(\underset{<\tau}{t_n},\tau+\Delta\tau)=\tilde{p}_j(t_n,\tau)\,\tilde{p}_j(\tau,\tau+\Delta\tau)\right)$

$$= P_{\gamma_{\mathrm{init}}}(0)\, J_{\tilde{w}}(\gamma) \times \begin{cases} \textbf{no jumps} \\ \tilde{p}_j(\tau,\tau+\Delta\tau) = 1-\Delta\tau\,\lambda_j(\tau)+O((\Delta\tau)^2) \\ \textbf{one jump} \\ \tilde{p}_j(\tau,t_f)\,w_{j\to k}(t_f)\,\tilde{p}(t_f,\tau+\Delta\tau) \\ \quad \text{“}\ w_{j\to k}(\tau)+O(\Delta\tau) \\ \text{for some } t_f\in(\tau,\tau+\Delta\tau) \text{ and } k\neq j \\ \textbf{more than one jump} \\ \text{higher orders in } \Delta\tau \end{cases}$$

$\parallel$

(probability density for γ)





We recover the definition of the process $\to$ proof of (3), (5)

§ expectation values and their relations

▷ **state quantity** $\hat{f}$ → takes value f_j in a (microscopic) state $j=1,\cdots,\Omega$

probability distribution $\mathbb{P}=(p_j)_{j=1,\cdots,\Omega}$

(1) $$\langle\hat{f}\rangle_{\mathbb{P}}=\sum_{j=1}^{\Omega}p_j f_j$$

▷ **jump quantity** $\hat{g}$ → takes value $g_{j\to k}$ when a jump $j\to k$ takes place

probability distribution $\mathbb{P}$, transition rates $\mathbb{W}=(w_{j\to k})_{j,k=1,\cdots,\Omega \ (j\neq k)}$

(2) $$\langle\hat{g}\rangle_{\mathbb{P},\mathbb{W}}=\sum_{\substack{j,k=1\\(j\neq k)}}^{\Omega}p_j w_{j\to k}\, g_{j\to k}$$ ← (the expectation value per unit time)

NOT a standard average!!

▷ **path quantity** $\hat{\hat{F}}$ → takes value $F(\gamma)$ in a path γ

(3) $$\langle\!\langle\hat{\hat{F}}\rangle\!\rangle_{\mathbb{P}(0),\tilde{\mathbb{W}}}=\int\mathcal{D}\gamma\, p_{\gamma_{\text{init}}}(0)\,\mathcal{J}_{\tilde{\mathbb{W}}}(\gamma)\,F(\gamma).$$

☐ Time-dependent state quantity $\hat{f}(t)$ → takes value $f_j(t)$

corresponding path quantity $\hat{\hat{f}}(t)$

takes value (1) $f(t,\gamma) = f_{\gamma(t)} = \sum_{m=0}^{n} f_{j_m}(t)\,\chi[t \in (t_m, t_{m+1}]]$

in path $\gamma = (j_0, \cdots, j_n; t_1, \cdots, t_n)$ $\quad t_0 = 0,\ t_{n+1} = \tau$

then (2) $\boxed{\langle\langle \hat{\hat{f}}(t) \rangle\rangle_{\mathbb{P}(0),\tilde{\omega}} = \langle \hat{f}(t) \rangle_{\mathbb{P}(t)}} = \sum_{j=1}^{\Omega} p_j(t) f_j(t)$

P46

integrated quantity (3) $\hat{\hat{F}} = \int_0^\tau dt\, \hat{\hat{f}}(t)$

takes value (4) $F(\gamma) = \int_0^\tau dt\, f(t,\gamma) = \sum_{m=0}^{n} \int_{t_m}^{t_{m+1}} dt\, f_{j_m}(t)$

thus

(5) $\langle\langle \hat{\hat{F}} \rangle\rangle_{\mathbb{P}(0),\tilde{\omega}} = \int_0^\tau dt\, \langle \hat{f}(t) \rangle_{\mathbb{P}(t)}$

[iii] time-dependent jump quantity $\hat{g}(t)$ → takes value $g_{j\to k}(t)$

corresponding path quantity $\hat{\hat{g}}(t)$

takes value (1) $g(t,\gamma) = \sum_{m=1}^{n} g_{j_{m-1}\to j_m}(t_m)\,\delta(t-t_m)$

in path $\gamma = (j_0, \dots, j_n; t_1, \dots, t_n)$ $\quad t_0 = 0,\ t_{n+1} = \tau$

then (2) $\boxed{\langle\langle \hat{\hat{g}}(t) \rangle\rangle_{\mathbb{P}(0),\tilde{\omega}} = \langle \hat{g}(t) \rangle_{\mathbb{P}(t),\omega(t)}} = \sum_{\substack{j,k=1 \\ (j\neq k)}}^{\Omega} p_j(t)\,\omega_{j\to k}(t)\,g_{j\to k}(t)$

P46

integrated quantity (3) $\hat{\hat{G}} = \int_0^\tau dt\ \hat{\hat{g}}(t)$

takes value (4) $G(\gamma) = \int_0^\tau dt\, g(t,\gamma) = \sum_{m=1}^{n} g_{j_{m-1}\to j_m}(t_m)$

thus

(5) $\langle\langle \hat{\hat{G}} \rangle\rangle_{\mathbb{P}(0),\tilde{\omega}} = \int_0^\tau dt\ \langle \hat{g}(t) \rangle_{\mathbb{P}(t),\omega(t)}$

derivation of P44-(2) and P45-(2)

$$f(t,\gamma) = f_{\gamma(t)} = \sum_{j=1}^{m} f_j(t)\,\delta_{\gamma(t),j}$$

since $\langle\langle \delta_{\gamma(t),j} \rangle\rangle_{\mathbb{P}(0),\tilde{\omega}} = P_j(t)$ $\Leftarrow$ p41-(5)

$$\therefore \langle\langle \hat{f}(t) \rangle\rangle_{\mathbb{P}(0),\tilde{\omega}} = \langle \hat{f}(t) \rangle_{\mathbb{P}(t)}$$

$$\int_t^{t+\Delta t} ds \langle\langle \hat{g}(s) \rangle\rangle_{\mathbb{P}(0),\tilde{\omega}} = \sum_{\substack{j,k \\ (j\neq k)}} g_{j\to k}(t)\,\mathrm{Prob}(\text{there is a jump } j\to k \text{ within } t\sim t+\Delta t) + O((\Delta t)^2)$$

$$= \sum_{j\neq k} \Delta t\, P_j(t)\, \omega_{j\to k}(t)\, g_{j\to k}(t) + O((\Delta t)^2)$$

$$= \Delta t \langle \hat{g}(t) \rangle_{\mathbb{P}(t),\omega(t)} + O((\Delta t)^2)$$

$$\therefore \langle\langle \hat{g}(t) \rangle\rangle_{\mathbb{P}(0),\tilde{\omega}} = \langle \hat{g}(t) \rangle_{\mathbb{P}(t),\omega(t)}$$

//

§ abstract fluctuation theorems for Markov jump processes

looks complicated BUT everything is formal and trivial

a Markov jump process with $\tilde{\omega} = (\omega_{k\to j}(t))_{k\neq j,\, t\in[0,\tau]}$

assume for any $j\neq k$ and $t\in[0,\tau]$ that (1) $\omega_{k\to j}(t)\neq 0 \Leftrightarrow \omega_{j\to k}(t)\neq 0$

the only assumption!

"entropy production" → formal definition

(2) $\boxed{\Theta^{\tilde{\omega}}_{k\to j}(t) := \log\dfrac{\omega_{k\to j}(t)}{\omega_{j\to k}(t)}}$ if $\omega_{k\to j}(t)\neq 0$ (set $\Theta^{\tilde{\omega}}_{k\to j}(t)=0$ otherwise)

we thus have

(3) $\omega_{k\to j}(t)\, e^{-\Theta^{\tilde{\omega}}_{k\to j}(t)} = \omega_{j\to k}(t)$ and (4) $\Theta^{\tilde{\omega}}_{j\to k}(t) = -\Theta^{\tilde{\omega}}_{k\to j}(t)$

for a path $\gamma = (j_0,\dots,j_n; t_1,\dots,t_n)$

(5) $\boxed{\Theta^{\tilde{\omega}}(\gamma) := \sum_{m=1}^{n} \Theta^{\tilde{\omega}}_{j_{m-1}\to j_m}(t_m)}$

total entropy production along the path γ

(path quantity

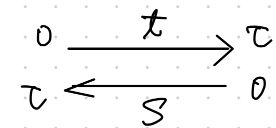
(6) $\hat{\Theta}^{\tilde{\omega}} = \int_0^\tau dt\, \hat{\Theta}^{\tilde{\omega}}(t)$)

time-reveresed Markov jump process

transition rates $\tilde{\omega}^\dagger = (\omega^\dagger_{k\to j}(s))_{k\neq j,\, s\in[0,\tau]}$ with (1) $\omega^\dagger_{k\to j}(s) = \omega_{k\to j}(\tau - s)$

$\tau - s = t$

(reversed time (2) $s = \tau - t$)

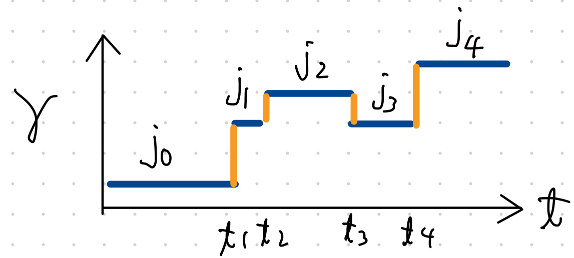


escape rate (3) $\lambda^\dagger_k(s) = \sum_{k(\neq j)} \omega^\dagger_{k\to j}(s) = \lambda_k(\tau - s)$

(4) $\tilde{\omega} = \tilde{\omega}^\dagger$ if $\tilde{\omega}$ is time-independent

time-reveresed path

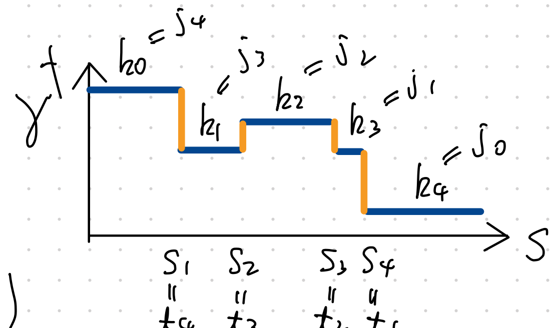
(5) $\gamma = (j_0, \cdots, j_n; t_1, \cdots, t_n)$



(6) $\gamma^\dagger = (j_n, \cdots, j_0; \tau - t_n, \cdots, \tau - t_1)$

$= (k_0, \cdots, k_n; s_1, \cdots, s_n)$

(7) $k_m = j_{n-m}$ (8) $s_m = \tau - t_{n-m+1}$



p47-(4) → (9) $\Theta^{\tilde{\omega}}(\gamma) = -\Theta^{\tilde{\omega}^\dagger}(\gamma^\dagger)$

(iii) detailed fluctuation theorem

P40-(4) P47-(2),(5)

(1) $$\mathcal{J}_{\tilde{\omega}}(\gamma)\, e^{-\Theta^{\tilde{\omega}}(\gamma)} = \prod_{m=0}^{n} \tilde{p}^{\tilde{\omega}}_{j_m}(t_m, t_{m+1}) \prod_{m=1}^{n} \omega_{j_{m-1}\to j_m}(t_m) \prod_{m=1}^{n} \frac{\omega_{j_m\to j_{m-1}}(t_m)}{\omega_{j_{m-1}\to j_m}(t_m)}$$

$$= \prod_{m=0}^{n} \tilde{p}^{\tilde{\omega}}_{j_m}(t_m, t_{m+1}) \prod_{m=1}^{n} \omega_{j_m\to j_{m-1}}(t_m)$$

(2) $$\tilde{p}^{\tilde{\omega}}_{j_m}(t_m, t_{m+1}) = \exp\left(-\int_{t_m}^{t_{m+1}} dt\, \lambda_{j_m}(t)\right) = \exp\left(-\int_{s_{n-m}}^{s_{n-m+1}} ds\, \lambda^{\dagger}_{k_{n-m}}(s)\right) = \tilde{p}^{\tilde{\omega}^{\dagger}}_{k_{n-m}}(s_{n-m}, s_{n-m+1})$$

(3) $$\omega_{j_m\to j_{m-1}}(t_m) = \omega^{\dagger}_{k_{n-m}\to k_{n-m+1}}(s_{n-m+1})$$

$$= \prod_{m=0}^{n} \tilde{p}^{\tilde{\omega}^{\dagger}}_{k_{n-m}}(s_{n-m}, s_{n-m+1}) \prod_{m=1}^{n} \omega^{\dagger}_{k_{n-m}\to k_{n-m+1}}(s_{n-m+1})$$

$$= \prod_{m=0}^{n} \tilde{p}^{\tilde{\omega}^{\dagger}}_{k_m}(s_m, s_{m+1}) \prod_{m=1}^{n} \omega^{\dagger}_{k_{m-1}\to k_m}(s_m) = \mathcal{J}_{\tilde{\omega}^{\dagger}}(\gamma^{\dagger})$$

(4) $$\boxed{\mathcal{J}_{\tilde{\omega}}(\gamma)\, e^{-\Theta^{\tilde{\omega}}(\gamma)} = \mathcal{J}_{\tilde{\omega}^{\dagger}}(\gamma^{\dagger})}$$

P47-(3)

(5) $$\omega_{k\to j}(t)\, e^{-\theta_{k\to j}(t)} = \omega_{j\to k}(t)$$

(iii) integrated fluctuation theorem

$\begin{cases} \mathbb{P}(0) \text{ arbitrary initial probability distribution with } P_j(0) \neq 0 \text{ for } {}^\forall j \\ \mathbb{q} = (q_j)_{j=1,\dots,\Omega} \text{ arbitrary probability distribution with } q_j \neq 0 \text{ for } {}^\forall j \end{cases}$

(1)

$$\left\langle\!\left\langle \exp\left[-\widehat{H}^{\tilde{\omega}} - \log P_{\gamma_{init}}(0) + \log q_{\gamma_{fin}}\right] \right\rangle\!\right\rangle_{\mathbb{P}(0),\tilde{\omega}}$$

$$= \int D\gamma\, P_{\gamma_{init}}(0)\, J_{\tilde{\omega}}(\gamma)\, e^{-H^{\tilde{\omega}}(\gamma)} \frac{1}{P_{\gamma_{init}}(0)}\, q_{\gamma_{fin}}$$

$D\gamma = D\gamma^\dagger$

the path probability in the process $\tilde{\omega}^\dagger$ with initial distribution $\mathbb{q}$

$$= \int D\gamma^\dagger\, q_{\gamma^\dagger_{init}}\, J_{\tilde{\omega}^\dagger}(\gamma^\dagger) = \int D\gamma\, \boxed{q_{\gamma_{init}}\, J_{\tilde{\omega}^\dagger}(\gamma)} = 1$$

(2)

$$\boxed{\left\langle\!\left\langle \exp\left[-\widehat{H}^{\tilde{\omega}} - \log P_{\gamma_{init}}(0) + \log q_{\gamma_{fin}}\right] \right\rangle\!\right\rangle_{\mathbb{P}(0),\tilde{\omega}} = 1}$$

▷ fluctuation theorem

$$\begin{cases} \text{a Markov jump process such that } \tilde{\omega} = \tilde{\omega}^{\dagger} \text{ (e.g. time-independent process)} \\ \mathbb{P}(0) \text{ arbitrary initial probability distribution with } P_j(0) \neq 0 \text{ for } {}^{\forall} j \end{cases}$$

define (1) $\Psi(\gamma) := \Theta(\gamma) + \log P_{\gamma_{init}}(0) - \log P_{\gamma_{fin}}(0)$

(2) $P(s) := \langle\langle \delta(\hat{\Psi} - s) \rangle\rangle_{\mathbb{P}(0), \tilde{\omega}}$ prob. density that $\Psi(\gamma)$ is equal to s

(3) $$P(s) e^{-s} = \int D\gamma \, P_{\gamma_{init}}(0) J_{\tilde{\omega}}(\gamma) \, \delta(\Psi(\gamma) - s) e^{-\Psi(\gamma)}$$

$$= \int D\gamma \, P_{\gamma_{init}}(0) J_{\tilde{\omega}}(\gamma) e^{-\Theta(\gamma)} \frac{P_{\gamma_{fin}}(0)}{P_{\gamma_{init}}(0)} \delta(\Psi(\gamma) - s)$$

$$= \int D\gamma \, P_{\gamma_{fin}}(0) J_{\tilde{\omega}^{\dagger}}(\gamma^{\dagger}) \, \delta(\Psi(\gamma) - s)$$

$\left(\Psi(\gamma^{\dagger}) = -\Psi(\gamma) \right)$ p48-(9)

$$= \int D\gamma^{\dagger} \, P_{\gamma^{\dagger}_{init}}(0) J_{\tilde{\omega}}(\gamma^{\dagger}) \, \delta(-\Psi(\gamma^{\dagger}) - s) = P(-s)$$

(4) $$\boxed{P(s) = e^{s} P(-s)}$$

Incomplete references

Cournot’s principle

G. Shafer, “Why did Cournot’s principle disappear?”, presentation slides (2006).
http://www.brunodefinetti.it/Bibliografia/disappear.pdf

Fluctuation theorem in the path description

C. Maes, “The Fluctuation Theorem as a Gibbs Property”, J. Stat. Phys. **95**, 367 (1999).

J. L. Lebowitz and H. Spohn, “A Gallavotti–Cohen-Type Symmetry in the Large Deviation Functional for Stochastic Dynamics”, J. Stat. Phys. **95**, 333 (1999).

G. Crooks, “Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences”, Physical Review E, **60**, 2721 (1999).