Derivation of Markov Process from Path Entropy Maximization

Julian Lee
School of Systems Biomedical Science
and
Dept. of Bioinformatics and Life Science
Soongsil University

OUTLINE

- Maximum Entropy Principle
- Path Entropy Maximization
- Derivation of Markov process

Maximum Entropy Principle

- E.T. Jaynes (Phys. Rev. 106, 620(1957), Phys. Rev. 108, 171(1957)): Statistical Mechanics as a logical inference
- Maximize Gibbs-Shannon entropy $-\sum p_i \log p_i$ under given constraints : Most unbiased estimate

Boltzmann distribution

- Constraint: mean energy $\sum p_i E_i = \epsilon$
- Normalization: $\sum p_i = 1$
- In the absence of other information (equilibrium), the most unbiased estimate of the probability distribution is obtained by maximizing

$$-\sum_i p_i \log p_i + \beta(\sum_i p_i E_i - \epsilon) + \nu(\sum p_i - 1)$$

Boltzmann distribution

$$-\sum_i p_i \log p_i + \beta(\sum_i p_i E_i - \epsilon) + \nu(\sum p_i - 1)$$

$$\delta p_i : -\ln p_i - 1 + \beta E_i + \nu = 0$$
$$\delta \beta : \sum p_j E_j - \epsilon = 0$$
$$\delta \nu : \sum p_j - 1 = 0$$

Boltzmann distribution

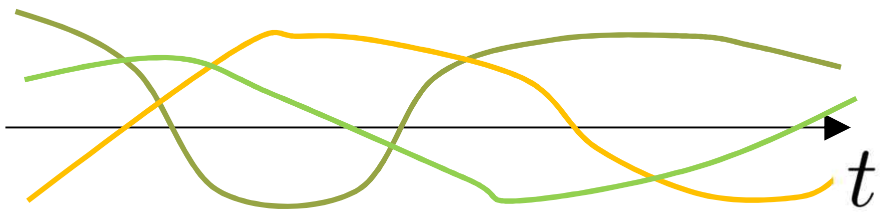
$$p_i = \frac{e^{-\beta E_i}}{\sum_j e^{-\beta E_j}}$$

With β determined by

$$\frac{\sum_k E_k e^{-\beta E_k}}{\sum_j e^{-\beta E_j}} = \epsilon$$

Path entropy maximization

❖Dynamical system: Obtain probability distribution $P(C)$ for path C



❖Maximize the path entropy $-\sum_C P(C) \ln P(C)$ under appropriate constraints

$\sum_C P(C) A^{(\alpha)}(C) = A_0^{(\alpha)}$ and $\sum_C P(C) = 1$

Path entropy maximization

- Jaynes ("Macroscopic prediction", in Complex Systems Operational Approaches in Neurobiology, Physics, and Computers, edited by H. Haken (Springer-Verlag, Berlin, 1985)): "Maximum Caliber principle"
- Filyukov and Karpov (J. Eng. Phys. 13, 624, 1967; 13, 798, 1967)
- Filyukov (Eng. Phys. Thermophys. 14, 814, 1968)

Discrete time (Filyukov et al.)

- $P(C) = p(i_0, i_1, \cdots i_T)$
- Path entropy: $H(T) = -\sum_{i_0, i_1, \cdots, i_T} p_{i_0 i_1 \cdots i_T} \log p_{i_0 i_1 \cdots i_T}$
- Stationary Markov process

$$p_C = p_{i_0} p_{i_0 \to i_1} p_{i_1 \to i_2} \cdots p_{i_{T-1} \to i_T}$$

$$H(T) = -\sum_i p_i \log p_i - T \sum_{i,j} p_i p_{i \to j} \log p_{i \to j}$$

- However, Markovian property itself can be derived from the property of the constraints!

Markov processes: Definitions

❖ n-point marginal probability

$$p(a_1, \cdots a_n; t) \equiv \sum_{i_0, i_1, \cdots i_{t-n}, j_1, j_2, \cdots j_{T-t}} p(i_0, i_1, \cdots i_{t-n}, a_1, \cdots a_n, j_1, j_2, \cdots j_{T-t})$$

❖ Transition probability

$$p(i_0, \cdots i_{t-1} \to i_t) \equiv \frac{p(i_0, \cdots i_t)}{p(i_0, \cdots i_{t-1})}$$

❖ n-th order Markov process

$$p(i_0, \cdots i_{t-1} \to i_t) = p(i_{t-n}, i_{t-n+1} \cdots i_{t-1} \to i_t; t) \equiv \frac{p(i_{t-n}, \cdots i_t; t)}{p(i_{t-n}, \cdots i_{t-1}; t)}$$

→ Transition probability depends only on previous n steps of history

Derivation of Markov processes

- n-th order Markov process follows if only up to (n+1)-point function is constrained
- General constraints:

$$\sum_C P(C)A^{(\alpha)}(C) = \sum_{\{i_0,i_1,\cdots i_T\}} p(i_0,i_1,\cdots i_T)A^{(\alpha)}(i_0,i_1,\cdots i_T) = A_0^{(\alpha)}$$

- One-point constraints: $A^{(\alpha)}(i_0,i_1,\cdots i_T) = \sum_{t=0}^{T} \varepsilon_{i_t}^{(\alpha)}$

$$F_0^{(\alpha)} \equiv \sum_{\{i_0,i_1,\cdots i_T\}} (\sum_{t=0}^{T} \varepsilon_{i_t}^{(\alpha)})p(i_0,i_1,\cdots i_T) - (T+1)E_0^{(\alpha)}$$

$$= \sum_{t=0}^{T}\sum_{i_t} \varepsilon_{i_t}^{(\alpha)} p(i_t;t) - (T+1)E_0^{(\alpha)} = 0 \quad (\alpha = 1,\cdots N_1)$$

Derivation of Markov processes

- Two-point constraints: $A^{(\alpha)}(i_0, i_1, \cdots i_T) = \sum_{t=0}^{T-1} J^{(\alpha)}_{i_t i_{t+1}}$

$$F_1^{(\alpha)} \equiv \sum_{\{i_0, i_1, \cdots i_T\}} \left(\sum_{t=0}^{T-1} J^{(\alpha)}_{i_t i_{t+1}}\right) p(i_0, i_1, \cdots i_T) - T J_0^{(\alpha)}$$

$$= \sum_{t=0}^{T-1} \sum_{i_t i_{t+1}} J^{(\alpha)}_{i_t i_{t+1}} p(i_t, i_{t+1}; t) - T J_0^{(\alpha)} = 0. \quad (\alpha = 1, \cdots N_2)$$

Derivation of Markov processes

❖ Take the variation of

$$-\sum_{\{i_0,i_1,\cdots i_T\}} p(i_0,i_1,\cdots i_T)\log p(i_0,i_1,\cdots i_T) - \sum_{\alpha=1}^{N_1}\beta_\alpha\left(\sum_{t=0}^{T}\sum_{i_t}\varepsilon_{i_t}^{(\alpha)}p(i_t;t) - (T+1)E_0^{(\alpha)}\right)$$

$$+\sum_{\gamma=1}^{N_2}\nu_\gamma\left(\sum_{t=0}^{T-1}\sum_{i_t i_{t+1}} J_{i_t i_{t+1}}^{(\gamma)} p(i_t,i_{t+1};t+1) - T J_0^{(\gamma)}\right) + (\rho+1)\left(\sum_{\{i_0,i_1,\cdots i_T\}} p(i_0,i_1,\cdots i_T) - 1\right)$$

$$\delta p: -\log p(i_0,i_1,\cdots i_T) - \sum_\alpha \beta_\alpha \sum_{t=0}^{T}\varepsilon_{i_t}^{(\alpha)} + \sum_\gamma \nu_\gamma \sum_{t=0}^{T-1} J_{i_t i_{t+1}}^{(\gamma)} + \rho = 0$$

$$p(i_0,i_1,\cdots i_T) \;=\; \exp\left(\rho - \sum_\alpha \beta_\alpha \sum_{t=0}^{T}\varepsilon_{i_t}^{(\alpha)} + \sum_\gamma \nu_\gamma \sum_{t=0}^{T-1} J_{i_t i_{t+1}}^{(\gamma)}\right)$$

Derivation of Markov processes

$$p(i_0, i_1, \cdots i_T) = \frac{v(i_0)G(i_0, i_1)G(i_1, i_2)\cdots G(i_{T-1}, i_T)v(i_T)}{\mathbf{v}^\dagger \mathbf{G}^T \mathbf{v}}$$

$$v(i) \equiv \exp\left(-\sum_\alpha \beta_\alpha \varepsilon_i^{(\alpha)}/2\right)$$

$$G(i,j) \equiv \exp\left(-\sum_\alpha \beta_\alpha \varepsilon_i^{(\alpha)}/2 + \sum_\gamma \nu_\gamma J_{ij}^{(\gamma)} - \sum_\alpha \beta_\alpha \varepsilon_j^{(\alpha)}/2\right)$$

Derivation of Markov processes

$$p(a_1, \cdots a_m; t) = \sum_{i_0, \cdots i_{t-m}, i_{t+1}, \cdots i_T} p(i_0, i_1, \cdots i_{t-m}, a_1, \cdots, a_m, i_{t+1}, \cdots, i_T)$$

$$= \frac{[\mathbf{v}^\dagger \mathbf{G}^{t-m+1}](a_1) G(a_1, a_2) G(a_2, a_3) \cdots G(a_{m-1}, a_m) [\mathbf{G}^{T-t} \mathbf{v}](a_m)}{\mathbf{v}^\dagger \mathbf{G}^T \mathbf{v}}$$

$$p(a_1, \cdots a_m \to a_{m+1}; t) = \frac{[\mathbf{v}^\dagger \mathbf{G}^{t-m}](a_1) G(a_1, a_2) \cdots G(a_m, a_{m+1}) [\mathbf{G}^{T-t} \mathbf{v}](a_{m+1})}{[\mathbf{v}^\dagger \mathbf{G}^{t-m}](a_1) G(a_1, a_2) \cdots G(a_{m-1}, a_m) [\mathbf{G}^{T-t+1} \mathbf{v}](a_m)}$$

$$= \frac{G(a_m, a_{m+1}) [\mathbf{G}^{T-t} \mathbf{v}](a_{m+1})}{[\mathbf{G}^{T-t+1} \mathbf{v}](a_m)} = p(a_m \to a_{m+1}; t).$$

Perron-Frobenius Theorem

(1) A positive matrix G has a positive real eigenvalue r, such that any other eigenvalue λ is strictly smaller than r in absolute value, $|\lambda| < r$.

(2) There is a left eigenvector $\mathbf{y}^\dagger = (y_1, \cdots y_N)$ for r with positive components. That is, $\mathbf{y}^\dagger \mathbf{G} = r\mathbf{y}^\dagger$ and $y_i > 0$ for all i. Similarly, there is a right eigenvector $\mathbf{z}$ with positive components, such that $\mathbf{Gz} = r\mathbf{z}$ and $z_i > 0$ for all i.

(3) Left and right eigenvectors with eigenvalue r are non-degenerate.

(4) $\lim_{T\to\infty} \frac{\mathbf{G}^T}{r^T} = \mathbf{z}\mathbf{y}^\dagger$

Time homogeneity

$$T - t \to \infty$$

$$p(a \to b; t) = \frac{G(a,b)[\mathbf{G}^{T-t}\mathbf{v}](b)}{[\mathbf{G}^{T-t+1}\mathbf{v}](a)} \to \frac{G(a,b)z(b)}{rz(a)}$$

$$t \to \infty$$

$$p(a;t) = \frac{[\mathbf{v}^{\dagger}\mathbf{G}^{t}](a)z(a)}{r^{t}\mathbf{v}^{\dagger}\mathbf{z}} \to y(a)z(a)$$

→Stationary Markov Process

Time homogenous master equation with an arbitrary initial distribution

- Initial condition is an additional information, which should also be implemented as a constraint

$$p(a; t = \tau) = \pi(a)$$

- Take variation of

$$-\sum_{\{i_0,i_1,\cdots i_T\}} p(i_0, i_1, \cdots i_T) \log p(i_0, i_1, \cdots i_T) - \sum_{\alpha=1}^{N_1} \beta_\alpha \left(\sum_{t=0}^{T} \sum_{i_t} \varepsilon_{i_t}^{(\alpha)} p(i_t; t) - (T+1) E_0^{(\alpha)} \right)$$

$$+ \sum_{\gamma=1}^{N_2} \nu_\gamma \left(\sum_{t=0}^{T-1} \sum_{i_t i_{t+1}} J_{i_t i_{t+1}}^{(\gamma)} p(i_t, i_{t+1}; t+1) - T J_0^{(\gamma)} \right) + (\rho + 1) \left(\sum_{\{i_0,i_1,\cdots i_T\}} p(i_0, i_1, \cdots i_T) - 1 \right)$$

$$+ \sum_a \lambda(a)(p(a; \tau) - \pi(a))$$

Time homogenous master equation

$$
\begin{aligned}
p(i_0, i_1, \cdots i_T) &= \exp(\rho + \lambda(i_\tau) - \beta \sum_{t=0}^{T} \varepsilon_{i_t} + \nu \sum_{t=0}^{T-1} J_{i_t i_{t+1}}) \\
&= \exp(\rho + \lambda(i_\tau)) v(i_0) G(i_0, i_1) G(i_1, i_2) \cdots G(i_{T-1}, i_T) v(i_T) \\
&= \frac{v(i_0)\pi(i_\tau) G(i_0, i_1) G(i_1, i_2) \cdots G(i_{T-1}, i_T) v(i_T)}{\sum_{j_0 \cdots j_T} v(j_0)\pi(j_\tau) G(j_0, j_1) G(j_1, j_2) \cdots G(j_{T-1}, j_T) v(j_T)}
\end{aligned}
$$

$\tau < t :$

$$
p(a_1, \cdots a_m \to a_{m+1}; t) = \boxed{\frac{G(a_m, a_{m+1})[\mathbf{G}^{T-t}\mathbf{v}](a_{m+1})}{[\mathbf{G}^{T-t+1}\mathbf{v}](a_m)}}
$$

$\tau \geq t :$

$$
p(a_1, \cdots a_m \to a_{m+1}; t) = \frac{G(a_m, a_{m+1}) \sum_a [\mathbf{G}^{\tau-t}](a_{m+1}, a)\pi(a)[\mathbf{G}^{T-\tau}\mathbf{v}](a)}{\sum_b [\mathbf{G}^{\tau-t+1}](a_m, b)\pi(b)[\mathbf{G}^{T-\tau}\mathbf{v}](b)}
$$

$$
= p(a_m \to a_{m+1}; t)
$$

→Markov process

Time homogenous master equation

- Again, for infinite duration of the constraints, the transition probability $p(a \to b; t)$ for $\tau \geq t$ is independent of time, and independent of the initial distribution $\pi(a)$.
- However, the state occupation probability $p(a; t)$ is time-dependent, with initial condition $p(a; t = \tau) = \pi(a)$
- Discrete time-homogeneous master equation:

$$p(a; t+1) = \sum_b p(b; t) p(b \to a)$$

Generalization

- ❖ Constraints on up to (n+1)-point probability leads to n-th point Markov process
- ❖ Condition for time-homogeneity ← Generalization of Perron-Frobenius theorem to higher rank tensor required.

Summary

- ❖ Path entropy maximization: Most unbiased estimated of the path probability under the given constraint
- ❖ In particular, no correlations exists except those given by the constraints => n-th order Markov process if only up to n-point function is constrained.
- ❖ http://arxiv.org/abs/1206.1416
- ❖ Collaboration with Steve Pressé (UCSF)