Alcune applicazioni del Principio della Massima Entropia

Maximum entropy description of a thermodynamic system in a stationary non equilibrium state

Marco Favretti

Dipartimento di Matematica Pura ed Applicata Università di Padova

June 5, 2009

Prologue

In the paper: L.M. Martyushev-V.D. Seleznev "*Maximum entropy production principle in physics, chemistry and biology*" Phys. Rep. 2006 reference is made to the paper A.A. Filyukov - V. Ya. Karpov " *Method of the most probable path of evolution in the theory of stationary irreversible processes*" Phys. Eng. J. 1967

... the paper did not attracted attention at their time but the method has much in common with approaches which were advanced much later[1] *and evoked great interest...*

[1] R.C. Dewar "*Maximum entropy production and the fluctuation theorem*" J. Phys. A (2005)

Starting from equilibrium statistical thermodynamics ....

- We consider a discretization of phase space Γ into n cells $\chi = \{1, \ldots, n\}$ (finite state space)
- We limit ourselves to some macroscopic properties of a generic trajectory of the dynamical system such as $Prob(\chi = i)$ proportional to time spent in cell i
- From experience: loss of information about microscopic dynamics does not impare reproductibility of macroscopic behaviour
- Problem: Assign $Prob(\chi = i)$ without relying on ergodicity hypothesis (top-down approach)

... to compute frequencies q in N independent trials

$\chi = \{1, \ldots, n\}$ discrete state space, assign p prior probability

$$\omega = (i_1, i_2, \ldots, i_N) \in \Omega_N = \chi^N \text{ discretized trajectory}$$

Hyp: χ_i i.i.d. random variables $\chi_i \sim p$

$$p(\omega) = p(i_1)p(i_2)\ldots p(i_N)$$

Define type of ω (frequency vector)

$$X(\omega) = (\frac{N_1(\omega)}{N}, \ldots, \frac{N_n(\omega)}{N}) =: (q_1, \ldots, q_n)$$

it gives macroscopic description of dynamics independent of initial conditions. Probability of a trajectory depends only on type of ω

$$p(\omega) = \prod_{i=1}^{n} p_i^{NX_i(\omega)} = e^{N \sum_i X_i(\omega) \ln p_i}$$

Enter entropy, relative entropy and cross entropy

To compute the probability of a trajectory

$$p(\omega) = \prod_{i=1}^{n} p_i^{NX_i(\omega)} = e^{N\sum_i X_i(\omega)\ln p_i}$$

use identity: cross entropy = entropy + relative entropy

$$-\sum_i q_i \ln p_i = -\sum_i q_i \ln q_i + \sum_i q_i \ln \frac{q_i}{p_i}$$

$$H(q;p) = H(q) + D(q\|p)$$

and get

$$p(\omega) = e^{-N[H(X(\omega))+D(X(\omega)\|p)]}$$

max for $X(\omega) = p$!

Let us compute probability of type q

$$Prob(X(\omega) = q) = \frac{N!}{\prod_{i=1}^{n}(Nq_i)!} e^{-N[H(q)+D(q\|p)]}$$

using Stirling approx $\ln n! \sim (N + \frac{1}{2}) \ln N - N + \frac{1}{2} \ln 2\pi$

$$Prob(X(\omega) = q) \sim \frac{e^{-ND(q\|p)}}{N^{\frac{n-1}{2}}} \frac{1}{\sqrt{\prod_i q_i}} \qquad \text{max for } q = p$$

Problems:

1. $Prob(X(\omega) = p) \to 0$ for $N \to \infty$
2. Granted that q with minimal relative entropy has a favorite status, how quickly are other p.d. q' ruled out?

Problem 1: Entropy and L.L.N. (Shannon Thm)

Theorem (weak L.L.N.)

For every $\varepsilon > 0$, if N is sufficiently large

$$Prob(\{\omega \in \Omega_N : |X(\omega) - p| < \varepsilon\}) \to 1$$

$$p(\omega) = e^{-N[H(q)+D(q\|p)]} \sim e^{-NH(p)}, \quad \text{for a.e.} \quad \omega$$

(ω typical sequences, asymptotic equipartition property).
If we introduce the joint entropy of N r.v. χ_i

$$H(\chi_1, \chi_2, \ldots, \chi_N) = \sum_{\omega \in \Omega_N} p(\omega) \ln p(\omega)$$

then $H(p)$ coincides with the entropy rate of the i.i.d. $\sim p$ process

$$H(p) = \lim_{N \to +\infty} \frac{1}{N} H(\chi_1, \chi_2, \ldots, \chi_N)$$

entropy rate $=$ thermodynamic limit of the entropy of χ

If constraints on empirical frequencies q are known

Theorem (l.l.n. with linear constraints, O.A. Vasicek, 1980)

$\forall \varepsilon > 0$ there exists $\delta_\varepsilon > 0$ such that $\forall \delta \in (0, \delta_\varepsilon]$

$$Prob(\{\omega \in \Omega_N : |X(\omega) - q| \leq \varepsilon \;\&\; |AX(\omega) - c| < \delta\}) \to 1$$

as $N \to \infty$, where q minimizes $D(q\|p)$ on the constraint.

Rem: if $p = 1/n$ then $H(q, 1/n) = H(q) + D(q\|1/n) = \ln n$
Problem : Granted that q with min rel entropy has a favorite status, how quickly are other p.d. q' ruled out?
Answer (Large deviation theory, E.C.T., Sanov Theorem)

$$Prob(q' : D(q'\|q) > \delta) = 1 - F(N\delta) = 1 - \int_0^{N\delta} p_{\chi^2_{n-k-1}}$$

Maximum enormously sharp for 'large' N.

Maximum entropy or minimum relative entropy principle

p : prior information, $F(q) \in C$: subsequent information

In the asymptotic N limit, probability distribution with max $H(q)$ [resp. min $D(q\|p)$] are enormously favoured with respect to all other p.d. satisfying constraints $F(q) \in C$

Max Ent is a consequence of large system dimensionality and scale separation

Credits:

L. Boltzmann $\sim$ 1870

J.W. Gibbs *Elementary Principles of Statistical mechanics* 1902

C. Shannon *A mathematical theory of communication* 1948

E.T. Jaynes *Information theory and Statistical Mechanics* 1957

Doing the same thing in the Markov chain setting

- $\chi\{1, \ldots, n\}$ state space,
- χ_i, $i \in \mathbb{N}$ not i.i.d. r.v. but satisfying Markov property

$$p(\omega) = Prob(\chi_1 = i_1, \ldots, \chi_N = i_N) = p(i_1)p(i_2|i_1)p(i_3|i_2 i_1)\ldots$$

$$= p(i_1)p(i_2|i_1)\ldots p(i_N|i_{N-1}) = p_i P_{i_1 i_2} \ldots P_{i_{N-1} i_N}$$

- P stochastic matrix of conditional probabilities
- Definition: a p.d. π is stationary for P if $P^t \pi = \pi$
- at time step k, χ_k is described by $\pi(k) = \pi P^k$

Law of large numbers for ergodic Markov processes

Let π be stationary p.d. for P. Consider stationary Markov process

$$p(\omega) = \pi(i_1) P_{i_1 i_2} \dots P_{i_{N-1} i_N} = \pi(i_1) \prod_{k,l=1}^{n} P_{kl}^{N_{kl}(\omega)}$$

if we define conditional frequencies $X_{kl}(\omega) = N_{kl}(\omega)/N_k(\omega)$ then

$$p(\omega) = \pi(i_1) e^{-N \sum_{kl} X_k(\omega) X_{kl} \ln P_{kl}}$$

As before: cross entropy = entropy + relative entropy

$$-\sum_{kl} q_k Q_{kl} \ln P_{kl} = -\sum_{kl} q_k Q_{kl} \ln Q_{kl} + \sum_{kl} q_k Q_{kl} \ln \frac{Q_{kl}}{P_{kl}}$$

$$H^q(Q; P) = H(q, Q) + D(Q \| P)$$

From l.l.n.

$$X_i(\omega) \sim \pi_i \qquad X_{kl}(\omega) = \frac{N_{kl}(\omega)}{N_k(\omega)} \sim P_{kl} \quad \forall k, l$$

hence, for a.e. ω

$$p(\omega) = \pi_i e^{-N[H(q,Q)+D(Q\|P)]} \sim e^{-NH(\pi,P)}$$

this is asymptotic equipartition property for M.C. expressed in terms of the entropy of the Markov chain

$$H(\pi, P) = -\sum_{kl} \pi_k P_{kl} \ln P_{kl}$$

As before the entropy of χ is the thermodynamic limit : for a.e. ω

$$\lim_N \frac{1}{N} H(\chi_1, \ldots, \chi_N) = \lim_N \frac{1}{N} \sum_{\omega \in \Omega_N} p(\omega) \ln p(\omega) = H(\pi, P)$$

Theorem

Let $P_{ij} > 0\ \forall i, j$. Then there exists a unique p.d. π such that $P^t\pi = \pi$ (stationary for P) and

$$\lim_{N\to\infty} (P^N)_{ij} = \pi_j$$

where $P^N = PP\dots P$ (N times). π is the equilibrium distribution

Strong ergodicity : setting $||\mu - \nu|| = \sum_i |\mu_i - \nu_i|$

$$\lim_{N\to\infty} ||\mu P^N - \pi|| = 0 \quad \forall\mu$$

asymptotic loss of memory of initial conditions
Remark: P determines equilibrium state π, converse implication is false.

M.E.P. for stationary Markov chains

- Let the equilibrium probability distribution π be given. Then we select stochastic matrix P which has π as stationary distribution, fulfills macroscopic constraints on the equilibrium state and has maximum entropy $H(\pi, P)$.
- This amount to choose the process that has an overwhelming number of possible realizations.
- As a consequence, we select the dynamics for the approach to equilibrium that has quickest loss of information about initial conditions
- Maximum entropy principle for Markov chain used in communication theory, reliability theory, seismic risk analysis

A model for a discrete system in S.NE.S.

- $\chi = \{1, \dots, n\}$ system, $\mathcal{A}$, $\mathcal{B}$ environments
  Hyp: the system is alternatively on contact with $\mathcal{A}$, $\mathcal{B}$.
- Introduce a time-dependent Markov chain with matrices $A > 0, B > 0$

$$P_t = \begin{cases} A & t = 2m, \\ B & t = 2m+1. \end{cases}$$

- Start the chain with given arbitrary p.d. π. Then (A^* = transposed matrix of A)

$$\pi(0) = \pi,\ \pi(1) = A^*\pi,\ \pi(2) = (AB)^*\pi,\ \pi(3) = (ABA)^*\pi, \dots$$

$$Prob(\chi_0 = i, \chi_1 = j, \chi_2 = k, \chi_3 = l) = \pi_i A_{ij} B_{jk} A_{kl}$$

- let $E = (E_1, \dots, E_n)$ be the energy of χ, $\mathbb{E}_\pi(E)$ is a macroscopic observable

imposing macroscopic constraints on A, B

$$A \cdot \mathbf{1} = B \cdot \mathbf{1} = \mathbf{1} \quad \text{(normalization of } A \text{ and } B\text{)}$$

Energy is conserved in microscopic transitions $i \to j$

$$\Delta E^{\chi}_{ij} = E_j - E_i = -\Delta E^{\mathcal{A}}$$

then average energy transfer in the contact with $\mathcal{A}$ is

$$(\Delta E^{\chi})_{av} = \sum_{ij} \pi_i A_{ij} \Delta E^{\chi}_{ij} = \mathbb{E}_{A^*\pi}(E) - \mathbb{E}_{\pi}(E) = -(\Delta E^{\mathcal{A}})_{av}$$

an energy $q = \dot{q}\tau \geq 0$ enters from $\mathcal{A}$ and is transferred on $\mathcal{B}$

$$\mathbb{E}_{A^*\pi}(E) - \mathbb{E}_{\pi}(E) = q, \quad \text{(specify energy inflow)}$$

after contact with $\mathcal{A}\&\mathcal{B}$ the system χ is unchanged

$$\mathbb{E}_{(AB)^*\pi}(E) - \mathbb{E}_{\pi}(E) = 0, \quad \text{(outflow = inflow)}$$

we set this last constraint in the form

$$(AB)^*\pi = \pi \quad \text{(stationarity of } \pi \text{ for } AB\text{)}$$

Computing the entropy

- If π is stationary for AB and A it is stationary also for B. This implies $A = B$, $q = 0$; the system is in a equilibrium state.
- If π is stationary for AB but not for A, the chain is weakly but not strongly ergodic. The system switches between

$$\pi, \quad A^*\pi, \quad \pi, \quad A^*\pi, \quad \pi, \ldots$$

- However its entropy is defined. We can compute the entropy of χ

$$p(\omega) = \pi_{i_1} A i_1 i_2 B i_2 i_3 A i_3 i_4 \ldots$$

due to the stationarity condition $(AB)^*\pi = \pi$

$$\mathcal{H} = \lim_N \frac{1}{N} H(\chi_1, \chi_2, \ldots, \chi_N) = \frac{1}{2}[H(\pi, A) + H(A^*\pi, B)]$$

We want to determine matrices A, B fulfilling constraints and maximizing $\mathcal{H}(A, B)$.

M.E. solution

Suppose that equilibrium p.d. is determined by additional constraint

$$\mathbb{E}_\pi(E) = e.$$

Then the maximum entropy solution for stochastic matrices A and B on the basis of the macroscopic constraints on e, average system energy, and q, average energy flux , is

$$A_{ij} = \pi'_j = \frac{e^{-\beta' E_j}}{Z(\beta')}, \qquad B_{ij} = \pi_j = \frac{e^{-\beta E_j}}{Z(\beta)}$$

where $\beta' = 1/T'$ and $\beta = 1/T$ are determined by

$$e + q = -\frac{\partial}{\partial \beta'} \ln Z(\beta'), \quad e = -\frac{\partial}{\partial \beta} \ln Z(\beta)$$

The system switches between p.d. π and π'

$$\pi(e) \xrightarrow{A} \pi'(e+q) \xrightarrow{B} \pi(e) \xrightarrow{A} \pi(e+q) \ldots$$

Entropy of χ

The entropy of χ is

$$\begin{aligned}\mathcal{H}(e,q) &= \frac{1}{2}[H(\pi,A)+H(\pi',B)] = \frac{1}{2}[H(\pi)+H(\pi')]\\ &= \frac{1}{2}[\ln Z(\beta)+\beta e+\ln Z(\beta')+\beta'(e+q)]\end{aligned}$$

If the energy flux q is small, we can expand $\mathcal{H}(e,q)$ in powers of q

$$\mathcal{H}(e,q) = \ln Z(\beta)+\beta e+\frac{1}{2}\frac{q}{T}+\mathcal{O}(q^2)$$

the entropy of the open system χ has a source and flux term. Moreover, up to $\mathcal{O}(q^2)$

$$q = \frac{\partial e}{\partial \beta}(\beta)d\beta = -\frac{\partial^2 \ln Z(\beta)}{\partial \beta^2}d\beta = -T^2 C_v d(\frac{1}{T}) = C_v dT.$$

The case of m synchronous fluxes

If the system is in a S.NE.S. with m fluxes q_α related to m macroscopic observables X_α other than the energy, and the equilibrium state is defined by m averages c_α, then the system switches between Gibbs states

$$\pi(c_1, \ldots, c_m) = \frac{e^{-\sum_i \beta_\alpha X_\alpha}}{Z(\beta_1, \ldots, \beta_m)}, \qquad \pi'(c_1 + q_1, \ldots, c_m + q_m)$$

satisfying Onsager' reciprocal relations

$$\frac{\partial c_\alpha}{\partial \beta_\gamma} = \frac{\partial c_\gamma}{\partial \beta_\alpha}$$

An example with 2 asynchronous fluxes

Suppose that the system returns (relaxes) to equilibrium with respect to observable X in a two step cycle

$$A \to B \qquad C \to D$$

while, with respect to observable Y relaxes to equilibrium in a full (four step) cycle

$$A \to B \to C \to D$$

Can we infer average values of slow quantity Y at intermediate states B, C from knowlege of average values of fast observable X ?

Constraints and entropy

$$A \cdot \mathbf{1} = B \cdot \mathbf{1} = C \cdot \mathbf{1} = D \cdot \mathbf{1} = \mathbf{1} \quad \text{(normalization)}$$

$$\mathbb{E}_{A^*\pi}(X) - \mathbb{E}_\pi(X) = q, \quad \text{(specify inflow of } X \text{ in } A\text{)}$$

$$\mathbb{E}_{(AB)^*\pi}(X) - \mathbb{E}_\pi(X) = 0, \quad \text{(specify outflow of } X \text{ in } AB\text{)}$$

$$\mathbb{E}_{(ABC)^*\pi}(X) - \mathbb{E}_\pi(X) = q, \quad \text{(specify inflow of } X \text{ in } ABC\text{)}$$

$$\mathbb{E}_{(AB)^*\pi}(Y) - \mathbb{E}_\pi(Y) = r, \quad \text{(specify inflow of } Y \text{ in } AB\text{)}$$

$$(ABCD)^*\pi = \pi, \quad \text{(stationarity of } \pi \text{ for } ABCD\text{)}.$$

$$\mathcal{H} = \frac{1}{4}[H(\pi, A) + H(A^*\pi, B) + H((AB)^*\pi, C) + H((ABC)^*\pi, D)]$$

M.E. solution

If equilibrium state is defined by $e = \mathbb{E}_\pi(X)$ and $u = \mathbb{E}_\pi(Y)$ then system switches between three Gibbsian states

$$\pi(e,u) \xrightarrow{A} \pi(e+q) \xrightarrow{B} \pi(e+q,u+r) \xrightarrow{C} \pi(e+q) \xrightarrow{D} \pi(e,u)$$

Then, the inferred average value of Y at intermediate states is

$$(Y)_{av} = \mathbb{E}_{\pi(e+q)}[Y]$$

Also

$$\frac{d(Y)_{av}}{dq}dq = -\frac{Cov_{\pi(e+q)}(X,Y)}{Cov_{\pi(e+q)}(X,X)}dq$$

Entropy production in closed system $\mathcal{A} \cup \chi \cup \mathcal{B}$

Let $\mathcal{A}$ and $\mathcal{B}$ be thermostats at temperatures $T_A = T' > T = T_B$. The entropy production in a $\pi \xrightarrow{A} \pi' \xrightarrow{B} \pi$ cycle is (χ is unchanged by AB)

$$S_{pr} = dS_A + dS_B = \frac{-q}{T'} + \frac{q}{T} = q(\beta - \beta')$$

It turns out that the entropy production of the overall, closed system $\mathcal{A} \cup \chi \cup \mathcal{B}$ is equal to the divergence (symmetrized relative entropy) between the probability distribution describing the two states assumed by χ

$$S_{pr}(e, q) = q(\beta - \beta') = [D(\pi \| \pi') + D(\pi' \| \pi)] = \Delta(\pi, \pi') \geq 0.$$

interpretation of divergence Δ

Let $\chi : \Omega \to \{1, \ldots, n\}$ be the r.v. describing the microscopic state. Consider hypotheses: χ is described by π [resp. by π']. Given observation $\chi = i$, by Bayes' Thm

$$P(\pi|i) = P(\pi)\frac{P(i|\pi)}{P(i)} = P(\pi)\frac{\pi_i}{P(i)}$$

$$\ln\frac{P(\pi|i)}{P(\pi'|i)} - \ln\frac{P(\pi)}{P(\pi')} = \ln\frac{\pi_i}{\pi'_i} \quad \text{gain of information in } i$$

$D(\pi\|\pi')$ is the average gain of information to discriminate between π and π' when χ is described by π

$$D(\pi\|\pi') = \sum_{i=1}^{n} \pi_i \ln\frac{\pi_i}{\pi'_i} \geq 0$$

symmetrize

$$\Delta(\pi;\pi') = D(\pi\|\pi') + D(\pi'\|\pi) = \sum_{i=1}^{n} (\pi_i - \pi'_i)\ln\frac{\pi_i}{\pi'_i} = S_{pr}$$

When a system is in a stationary nonequilibrium state described by a flux of energy q between two thermostats at different temperatures $T'(e+q) > T(e)$, the entropy production

$$S_{pr} = dS_A + dS_B = \frac{-q}{T'} + \frac{q}{T} = q\frac{T' - T}{TT'} = \Delta$$

is a measure of the information available through observations to determine in which state π or π' the system is (divergence between π and π')

$$S_{pr} = \Delta \text{ high } \rightarrow \quad \text{easy to discriminate}$$

$$S_{pr} = \Delta \text{ low } \rightarrow \quad \text{difficult to discriminate}$$