35 Large deviation approach to dynamical systems

35.1 Level 1 Large deviation: review

We have already discussed large deviation theory in **31.8** for Bernoulli samples, but here we wish to be a bit more systematic.

Our starting point is the law of large numbers. Consider a chaotic endomorphism f. If it is chaotic enough (e.g., topologically mixing), the law of large numbers holds for the time average:

$$\lim_{N\to\infty} P\left(\frac{1}{N}\sum_{k=0}^{N-1} f^k(x) \sim \langle f\rangle\right) = 1, \tag{35.1}$$

where $\langle f\rangle$ is the average. Here, $f^k(x)$ may be denoted as $x(k,\omega)$ as well (i.e., the trajectory (specified by) ω observed at time k). The long-time average $\langle f\rangle$ is the true expectation value (ergodicity).

Now, what happens if time is not long enough? The average should deviate (fluctuate) from the true average. How? This is expressed as the large-deviation principle

$$P\left(\frac{1}{N}\sum_{k=0}^{N-1} f^k(x) \sim x\right) \approx e^{-NI(x)}, \tag{35.2}$$

where $I(x)$ is called the rate function or the large deviation function. It is a convex function with the unique minimum at $x = \langle f\rangle$. Thus, (35.2) includes the law of large numbers.

35.2 Level 2 Large deviation: introduction

Sanov's theorem **31.8** we already discussed for simple cases is actually a 'level 2' theory.

We can study the invariant measure empirically, taking statistics, so there must be a corresponding law of large numbers

$$\lim_{N\to\infty} P\left(\frac{1}{N}\sum_{k=0}^{N-1} \delta(x(k,\omega) - \cdot) \sim \frac{d\mu(\cdot)}{dm(\cdot)}\right) = 1. \tag{35.3}$$

Here, m is the basic sampling measure (for the real space, it is usually the Lebesgue measure λ, so m and λ may be used interchangeably).

There must be a corresponding large deviation principle. The LD for expectation values is called the level 1 LD, and this one the level 2.

$$P\left(\frac{1}{N}\sum_{k=0}^{N-1}\delta(x(k,\omega)-\cdot)\sim\frac{d\mu(\cdot)}{dm(\cdot)}\right)\approx e^{-NI^{(2)}(\mu)}. \tag{35.4}$$

Here $I^{(2)}$ is the rate function(al) which is a convex functional of the measures with the unique minimum at the invariant measure compatible with the sampling measure. As is clear from the definition, μ must be absolutely continuous with respect to m.

35.3 Gärtner-Ellis theorem: level 1

We make a 'partition function' (generator)

$$Z(t)=\left\langle\exp\left(t\sum_{k=0}^{N-1}x(k,\omega)\right)\right\rangle, \tag{35.5}$$

where $\langle\ \rangle$ implies the average over the sampling measure m. Let us use (35.2) to compute this average:

$$Z(t)=\int dm(\omega)\int dx\,\delta\left(\frac{1}{N}\sum_{k=0}^{N-1}x(k,\omega)\sim x\right)e^{Ntx}=\int dx\,e^{N[tx-I(x)]}=e^{Nq(t)}, \tag{35.6}$$

where $q(t)$ is the $(-)$ free energy. Since N is very large, the integral is dominated by the peak value of the integrand, i.e., 'max'$_x[tx-I(x)]$. Therefore, we obtain

$$q(t)=\sup_x[tx-I(x)]. \tag{35.7}$$

Since $I(x)$ is convex, this is a proper Legendre transformation, and $q(t)$ is also a convex function. Thus,

$$I(x)=\sup_t[tx-q(t)]. \tag{35.8}$$

That is, 'entropy' (i.e., $-\log$ 'probability') may be obtained from 'free energy' through a Legendre transformation: q is much easier to compute than I, but I dictates what you can observe.

35.4 Gärtner-Ellis theorem: level 2

There must be a level 2 version of 'statistical mechanics.' We simply mimic the level 1 **35.3**. Make the partition function(al):

$$Z(\phi) = \int dm(\omega) \int \delta\mu\, \delta\left(\frac{1}{N}\sum_{k=0}^{N-1} \delta(x(k,\omega) - \cdot) \sim \frac{d\mu(\cdot)}{dm(\cdot)}\right) e^{N\int d\mu(\cdot)\,\phi(\cdot)} \tag{35.9}$$

$$= \int \delta\mu\, e^{N[\int d\mu(\cdot)\,\phi(\cdot) - I^{(2)}(\mu)]} = e^{Nq(\phi)}, \tag{35.10}$$

where $q(\phi)$ is the $(-)$ free energy functional, and $\int \delta\mu$ is a functional integral. Since N is very large, the integral is dominated by the peak value of the integrand. Therefore, we obtain

$$q(\phi) = \sup_{\mu}\left[\int d\mu(\cdot)\,\phi(\cdot) - I^{(2)}(\mu)\right]. \tag{35.11}$$

Since $I^{(2)}(\mu)$ is convex, this is a proper Legendre transformation, and $q(\phi)$ is also a convex function(al). Thus,

$$I^{(2)}(\mu) = \sup_{\phi}\left[\int d\mu(\cdot)\,\phi(\cdot) - q(\phi)\right]. \tag{35.12}$$

That is, 'entropy' may be obtained from 'free energy' through a Legendre transformation: q is much easier to compute than I, but I dictates what you can observe just as at level 1.

35.5 Sanov's theorem revisited

Let us compute $I^{(2)}(\mu)$ when the sampling measure is m. Notice that the partition function (35.9) can be rewritten as

$$Z(\phi) = \int dm(\omega)\, e^{\sum_{k=0}^{N-1} \phi(x(k,\omega))}, \tag{35.13}$$

because μ is an empirical measure obtained from $\{x(k,\omega)\}_{k=0}^{N-1}$. Let us assume that $x(k,\omega)$ are statistically independent for different k (i.e., let us assume the dynamics is Bernoulli). Then,

$$Z(\phi) = \int dm \prod_{k=0}^{N-1} e^{\phi(x(k,\omega))} = z^N, \tag{35.14}$$

where

$$z(\phi) = \int dm\, e^{\phi(x)}, \tag{35.15}$$

Therefore, $q(\phi) = \log z(\phi)$ or

$$I^{(2)}(\mu) = \sup_{\phi} \left[\int d\mu(\cdot)\, \phi(\cdot) - \log z(\phi) \right]. \tag{35.16}$$

Let us compute ϕ that 'maximizes' [ ]. The functional derivative gives[398]:

$$\int d\mu(x)\, \delta(x-y) - \frac{1}{z(\phi)} \frac{\delta z}{\delta \phi(y)} = 0, \tag{35.17}$$

where

$$\frac{\delta z}{\delta \phi(y)} = \int dm\, \delta(x-y) e^{\phi(x)} \tag{35.18}$$

Therefore, we get

$$d\mu(y) = \frac{dm(y) e^{\phi(y)}}{z(\phi)} \quad \Rightarrow \quad \frac{d\mu}{dm} = \frac{e^{\phi(y)}}{z(\phi)}. \tag{35.19}$$

That is,

$$\phi(y) = \log \frac{d\mu}{dm} + \log z. \tag{35.20}$$

Introducing this into (35.16), we obtain

$$I^{(2)}(\mu) = \int d\mu(\cdot)\, \log \frac{d\mu}{dm}. \tag{35.21}$$

This is called Sanov's theorem, justifying the use of information theory to search for the most probable results.

Notice, however, the result is only for Bernoulli processes. We must relax this constraint.

35.6 Extension of Sanov-type formula for general level 2

Warning: I identify the actual trajectories in M and their counterpart symbol sequences. In short I identify as measurable spaces the actual dynamical system and its isomorphic symbolic dynamics.

The easiest way is to regard a chunk of a trajectory (consecutive n time points) as a single state. Thus,

$$I^{(2)}(\mu) = \int d\mu(\cdot)\, \log \frac{d\mu}{dm}. \tag{35.22}$$

[398] The basic measure is m. That is δ-function is with respect to measure m

reads (recall the notation $[x] = x_1 \cdots x_n$)

$$I^{(2)}(\mu) = \int d\mu([x]_n) \log \frac{d\mu}{dm}([x]_n). \tag{35.23}$$

The Radon-Nikodym derivative can be rewritten as follows, assuming n is sufficiently large:

$$\frac{d\mu}{dm}([x]_n) = \frac{\mu([x]_n)}{m([x]_n)}. \tag{35.24}$$

Using the conditional probability

$$m([x]_n) = m(x_1 \mid x_2 \cdots x_n)m(x_2 \cdots x_n) \tag{35.25}$$

therefore, in the large n limit we can write

$$m([x_1 \cdots x_n]) = \frac{m([x_1 x_2 \cdots x_n])}{m([x_2 \cdots x_n])} \frac{m([x_2 x_3 \cdots x_n])}{m([x_3 \cdots x_n])} \cdots \tag{35.26}$$

$$= \frac{m([x_1 x_2 \cdots x_n])}{m(f[x_1 x_2 \cdots x_n])} \frac{m([x_2 \cdots x_n])}{m(f[x_2 \cdots x_n])} \cdots. \tag{35.27}$$

Therefore,

$$\frac{1}{n} \log m([x_1 \cdots x_n]) = \frac{1}{n} \sum_{j=1}^{n-1} \log \frac{m([x_j \cdots x_n])}{m(f[x_j \cdots x_n])}. \tag{35.28}$$

Analogously we get

$$\frac{1}{n} \log \mu([x_1 \cdots x_n]) = \frac{1}{n} \sum_{j=1}^{n-1} \log \frac{\mu([x_j \cdots x_n])}{\mu(f[x_j \cdots x_n])}. \tag{35.29}$$

Thus, if we consider $(1/n)I^{(2)} \to I^{(3)}$ in the large n limit, we get

$$I^{(3)} = \int d\mu(x) \left[\log \frac{d\mu}{d\mu \circ f} - \log \frac{dm}{dm \circ f} \right]. \tag{35.30}$$

Recall (31.3). The first term is (−)KS-entropy. Notice that this is the large deviation function for trajectories (so it is called the level 3 rate function).

Although I said I would identify the real space and the symbol space, and although the identification is almost perfect (it is perfect for 1D endomorphism, so this immediately gives Rohlin's formula), the formula must be carefully interpreted. For a sequence $\{x_1, x_2, x_3, \cdots\}$, $f(\{x_1, x_2, x_3, \cdots\}) = \{x_2, x_3, \cdots\}$ (cf. the shift). Notice that one symbol disappears. This shortening of the cylinder set implies expansion. That is why the first term is $(-)$KS-entropy. The second term is the expansion rate; if we have a nice coding such as Markov partitions, then it is the sum of positive LCN. That is, (35.30) reads

$$I^{(3)} = \sum_{+} \langle \chi \rangle_\mu - h_\mu \geq 0 \tag{35.31}$$

Thus, Pesin's equality holds for the most probable state = equilibrium state.

35.7 Energy function

To do study statistical mechanic on the lattice we need an energy function $\varphi : \Sigma_n \to \mathbb{R}$. It is the 1/2 of the total interaction and the self energy of a spin. we impose the following constraint on ϕ. Let

$$\mathrm{var}_k \phi = \sup\{|\phi(x) - \phi(y)| \, : x_i = y_i \text{ for } |i| \leq k\}. \tag{35.32}$$

Our constraint is

$$\mathrm{var}_k \varphi \leq b\alpha^k, \tag{35.33}$$

where $b > 0$ and $\alpha \in (0, 1)$. Let us explicitly write

$$\mathcal{F}_A = \{\phi \,|\, \Sigma_A \to \mathbb{R}, \mathrm{var}_k \varphi \leq b\alpha^k, \forall k \in \mathbb{N}\}. \tag{35.34}$$

Here Σ_A is a Markov subshift with matrix A.

35.8 Gibbs measure

For any Hamiltonian $\phi \in \mathcal{F}_A$, there is a unique shift invariant measure μ_ϕ satisfying the following inequality for some positive constants C_1, C_2 and P (= free energy per spin)

$$C_1 \leq \frac{\mu\{y \, : y_i = x_i, \forall i \in \{0, 1, \cdots, n\}}{\exp(-Pm + \sum_{k=0}^{n-1} \phi(\sigma^k x))} \leq C_2, \tag{35.35}$$

35.9 Transfer operator

Consider the totality of the states on the right-half lattice S_A^+.[399] and $\phi \in C(\Sigma_A^+$.

[399] If you read the original math paper, there is a log discussion about how to justify considering of Σ^+ instead of S.

Define the transfer operator T_ϕ as

$$[T_\phi f](x) = \sum_{y \in \sigma^{-1}x} e^\phi(y) f(y). \tag{35.36}$$

Notice that $y = x_* x_1 x_2 \cdots$.

35.10 Ruelle-Perron-Frobenius theorem

Let Σ_A be mixing and $\phi \in \mathcal{F}_A \cap C(\Sigma_A^+)$. There is a unique positive eigenvalue λ_ϕ of T_ϕ, and the Gibbs measure is obtained from the partition func tion and the normalization obtained from λ_ϕ

$$\mu([x]_n) \simeq \frac{1}{\lambda_\phi^n} \exp\left(\sum_{i=1}^n \phi(x_i)\right), \tag{35.37}$$

where $[x]_n = x_1 \cdots x_n$.

35.11 Variational principle for Gibbs measure

The Gibbs measure μ_ϕ is the unique measure satisfying the following variational principle:

$$s(\mu) + \int \phi d\mu = P(\phi). \tag{35.38}$$

where s is the entropy per spin.

The T-invariant measure satisfying the variational principle is called an equilibrium state wrt to T and ϕ.[400]

[400] Th 2.7