Mean-field Interacting Particle Systems: Limit Laws and Large Deviations

Rajesh Sundaresan and Sarath Yasodharan

Indian Institute of Science and Brown University

SIGMETRICS/PERFORMANCE 2022
10 June 2022

Outline

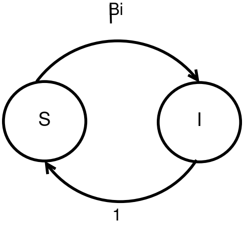
1. Model description and the mean-field limit (Rajesh)
2. Large deviation from the mean field limit: finite durations and the stationary regime (Sarath)
3. Two time-scale systems (Sarath)
4. Some interesting phenomena in infinite state space systems (Rajesh)

Section 1

Model description and the mean-field limit

A mean-field SIS epidemic model

- Interacting system with N individuals
- Each node's state space: $\mathcal{Z} = \{\mathrm{S}, \mathrm{I}\}$
- Transitions:



- Dynamics depends on the "mean field". Global interaction.
  $\mu_N(t) = i =$ fraction of nodes in infectious state
- Transition rate from S to I or I to S depends on the fraction of nodes in the infectious state. $\lambda_{S,I}(\mu_N(t)) = \beta i$ and $\lambda_{I,S}(\mu_N(t)) = 1$.

Reversible versus nonreversible dynamics

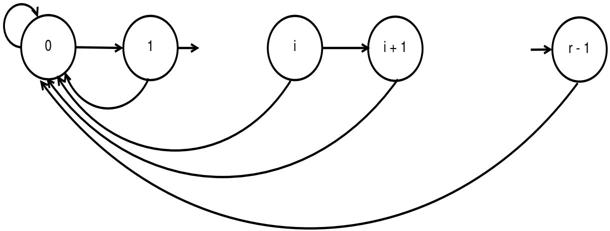
- (Reversible) Gibbsian system
  - Example: Heat bath dynamics
  - State space $\mathcal{Z} = \{0, 2, \ldots, r-1\}$
  - Configuration of the N particles $x = (x_1, \ldots, x_N)$. μ_N: empirical measure
  - $E(\mu_N)$: Energy of a configuration $x = (x_1, \ldots, x_N)$ with mean μ_N
  - An i to j transition takes μ_N to $\mu_N - \frac{1}{N}\delta_i + \frac{1}{N}\delta_j$

$$\lambda_{ij}(\mu_N) = \frac{e^{-NE(\mu_N)}}{e^{-NE(\mu_N - \frac{1}{N}\delta_i + \frac{1}{N}\delta_j)} + e^{-NE(\mu_N)}}$$

- In general, $\lambda_{ij}(\cdot)$ may result in nonreversible dynamics
- Weak interaction

Wireless Local Area Network (WLAN) interactions DCF 802.11 countdown and its CTMC caricature

- N particles accessing the common medium in a wireless LAN
- Each particle's state space: $\mathcal{Z} = \{0, 1, \cdots, r-1\}$
- Transitions:



  - State = # of transmission attempts for head-of-line packet
  - r: Maximum number of transmission attempts before discard

- Coupled dynamics: Transition rate for success or failure depends on empirical distribution $\mu_N(t)$ of particles across states

Example transition rates

- Matrix of rates: $\Lambda(\cdot) = [\ \lambda_{i,j}(\xi)\ ]_{i,j\in\mathcal{Z}}$.
- Assume three states, $\mathcal{Z} = \{0, 1, 2\}$ or $r = 3$.
- Aggressiveness of the transmission $c = (c_0, c_1, c_2)$.
- Conventional wisdom, double the waiting time after every failure, $c_i = c_{i-1}/2$.
- For μ, the empirical measure of a configuration, the rate matrix is

$$\Lambda(\mu) = \begin{bmatrix} -(\cdot) & c_0(1 - e^{-\langle\mu,c\rangle}) & 0 \\ c_1 e^{-\langle\mu,c\rangle} & -(\cdot) & c_1(1 - e^{-\langle\mu,c\rangle}) \\ c_2 e^{-\langle\mu,c\rangle} & 0 & -(\cdot) \end{bmatrix}.$$

- "Activity" coefficient $a = \langle\mu, c\rangle$.
  Probability of no activity $= e^{-a}$.

Mean-field interaction and dynamics

- Configuration $X^N(t) = (x_1(t), \ldots x_N(t))$.
- Empirical measure $\mu_N(t)$: Fraction of particles in each state
- A particle transits from state i to state j at time t with rate $\lambda_{i,j}(\mu_N(t))$

Studying the time-evolutions

- Tag a particle, say n_1. Study $X_{n_1}^{(N)}(\cdot)$. Marginal at n_1.

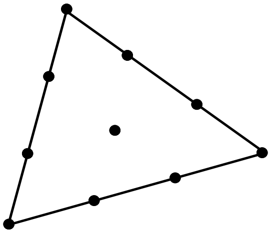
- Tag two particles, say n_1, n_2. Study $(X_{n_1}^{(N)}(\cdot), X_{n_2}^{(N)}(\cdot))$, marginals at n_1, n_2.

- Study $\mu_N(\cdot)$.

The Markov processes, big and small

- $(X_n^{(N)}(\cdot),\ 1 \leq n \leq N)$, the trajectory of all the n nodes, is Markov

- Study $\mu_N(\cdot)$ instead, also a Markov process
  Its state space size is the set of empirical probability measures on N particles with state space $\mathcal{Z}$.



- Then try to draw conclusions on the original process.

The smaller Markov process $\mu_N(\cdot)$

- A Markov process with state space being the set of empirical measures of N nodes.
- This is a measure-valued flow across time.
- The transition $\xi \rightsquigarrow \xi + \frac{1}{N}e_j - \frac{1}{N}e_i$ occurs at rate $N\xi(i)\lambda_{i,j}(\xi)$.
- For large N, changes are small, $O(1/N)$, at higher rates, $O(N)$. Individuals are collectively just about strong enough to influence the evolution of the measure-valued flow.
- Fluid limit : μ_N converges to a deterministic limit given by an ODE.

The conditional expected drift in μ_N

- Recall $\Lambda(\cdot) = [\ \lambda_{i,j}(\cdot)\ ]$ without diagonal entries. Then

$$\lim_{h \downarrow 0} \frac{1}{h}\, \mathbb{E}\left[\mu_N(t+h) - \mu_N(t) \mid \mu_N(t) = \xi\right] = \Lambda(\xi)^T\ \xi$$

with suitably defined diagonal entries.

An interpretation

- The rate of change in the kth component is made up of increase

$$\sum_{i:i\neq k} (N\xi_i) \cdot \lambda_{i,k}(\xi) \cdot (+1/N)$$

- and decrease

$$(N\xi_k) \sum_{i:i\neq k} \lambda_{k,i}(\xi)(-1/N).$$

- Put these together:

$$\sum_{i:i\neq k} \xi_i \lambda_{i,k}(\xi) - \xi_k \sum_{i:i\neq k} \lambda_{k,i}(\xi) = \sum_i \xi_i \lambda_{i,k}(\xi) = (\Lambda(\xi)^T \xi)_k.$$

The conditional expected drift in μ_N

- Recall $\Lambda(\cdot) = [\ \lambda_{i,j}(\cdot)\ ]$ without diagonal entries. Then

$$\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}\left[\mu_N(t+h) - \mu_N(t) \mid \mu_N(t) = \xi\right] = \Lambda(\xi)^T\ \xi$$

  with suitably defined diagonal entries.

- Anticipate that $\mu_N(\cdot)$ will solve (in the large N limit)

$$\begin{aligned} \dot{\mu}(t) &= \Lambda(\mu(t))^T\ \mu(t), \quad t \geq 0 \qquad \text{[McKean-Vlasov equation]} \\ \mu(0) &= \nu \end{aligned}$$

- Nonlinear ODE.

ODE preliminaries

$$\begin{aligned} \dot{\mu}(t) &= F(\mu(t)), \quad t \geq 0 \\ \mu(0) &= \nu \end{aligned}$$

- $C([0,T],\mathbb{R}^r)$: space of continuous functions from $[0,T]$ to $\mathbb{R}^r$.
- Can define a norm and a distance on this space:

$$\|\mu\| = \sup_{t\in[0,T]} \|\mu(t)\|$$
$$d_T(\mu,\xi) = \|\mu - \xi\|.$$

- $C([0,\infty),\mathbb{R}^r)$ with metric $d(\mu,\xi) = \sum_{T=1}^{\infty} 2^{-T}\left(d_T(\mu|_T,\xi|_T) \wedge 1\right)$.
- An ODE is well-posed if
  - For each $\nu \in \mathbb{R}^r$, the ODE has a unique solution $\mu(\cdot)$ on $[0,\infty)$
  - The mapping $\nu \mapsto \mu(\cdot) \in C([0,\infty),\mathbb{R}^r)$ is continuous.

Theorem

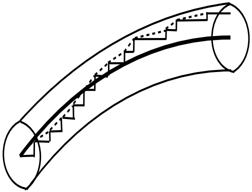
If F is Lipschitz, then the ODE is well-posed, and the solution can be written as $\mu(t) = \nu + \int_0^t F(\mu(s))\,ds$ for $t \in \mathbb{R}_+$.

Convergence in probability

- $\mu_N(\cdot)$ a sample path (random) while $\mu(\cdot)$ some deterministic or random path
- Fix T. View $\mu_N(\cdot)$ (interpolated) and $\mu(\cdot)$ as elements of $C([0,T],\mathcal{M}_1(\mathcal{Z}))$.
- We say $\mu_N(\cdot) \to \mu(\cdot)$ if for every $\varepsilon > 0$, we have

$$\Pr\{d_T(\mu_N(\cdot),\mu(\cdot)) > \varepsilon\} \to 0 \text{ as } N \to \infty$$

- This is the same as asking that the path $\mu_N(\cdot)$ remains within any ε-tube of $\mu(\cdot)$ with probability approaching 1 as $N \to \infty$.



A limit theorem

Theorem

Suppose that the initial empirical measure $\mu_N(0) \xrightarrow{p} \nu$*, where* ν *is deterministic.*

Assume each $\lambda_{i,j}(\cdot)$ *is Lipschitz in its argument. Let* $\mu(\cdot)$ *be the solution to the McKean-Vlasov dynamics with initial condition* $\mu(0) = \nu$*.*

Then $\mu_N(\cdot) \xrightarrow{p} \mu(\cdot)$*.*

Technicalities:

- Fix $T > 0$ and $\varepsilon > 0$. We will argue

$$\begin{aligned}\Pr\{d_T(\mu_N, \mu) > \varepsilon\} \quad \leq \quad & \Pr\{\|\mu_N(0) - \mu(0)\| > \varepsilon/(2e^{MT})\} \\ & + C_1 \exp\{-NT\overline{\lambda}h(\varepsilon/(C_2 Te^{MT}))\}\end{aligned}$$

  where M is the Lipschitz constant of the driving function, $\overline{\lambda}$ is the max of the transition rates, and $h(t) = (1+t)\ln(1+t) - t, \quad t > -1$.

Back to the individual nodes

- Let $\mu(\cdot)$ be the solution to the McKean-Vlasov dynamics

- Choose a node uniformly at random, and tag it.

  - $\mu_N(\cdot)$ is the distribution for the state of the tagged node at time t.
  - As $N \to \infty$, the limiting distribution is then $\mu(t)$

Joint evolution of tagged nodes

Theorem

Fix t, k. Tag k nodes at random.
Let $(X_n^{(N)}(0), 1 \leq n \leq N)$ be exchangeable and let $\mu_N(0) \xrightarrow{d} \nu$, a fixed limiting initial condition. Assume all transition rates are Lipschitz functions. Then

$$(X_{n_1}^{(N)}(t), \ldots, X_{n_k}^{(N)}(t)) \xrightarrow{d} (U_1, \ldots, U_k)$$

where $U_1, \ldots, U_k$ are iid with distribution $\mu(t)$.

- If the interaction is only through $\mu_N(t)$, and this converges to a deterministic $\mu(t)$, the transition rates are just $\lambda_{i,j}(\mu(t))$.
- Each of the k nodes is then executing a time-dependent Markov process with transition rate matrix $\Lambda(\mu(t))$.
- Asymptotically, no interaction (decoupling). The node trajectories are (asymptotically) iid (i.e., $\mu(t) \otimes \cdots \otimes \mu(t)$).

Stationary regime

- Interest in large time behaviour for a finite N system: $\lim_{t\to\infty} \mu_N(t)$.
  If N is large, we really want:

$$\lim_{N\to\infty} \left[\lim_{t\to\infty} \mu_N(t)\right].$$

- Idea: Try to predict where the system will settle from the following:

$$\lim_{t\to\infty} \left[\lim_{N\to\infty} \mu_N(t)\right] = \lim_{t\to\infty} \mu(t).$$

A fixed-point analysis

- Solve for the rest point of the dynamical system: $\dot{\mu}(t) = \Lambda(\mu(t))^T \mu(t)$, i.e., solve for ξ in

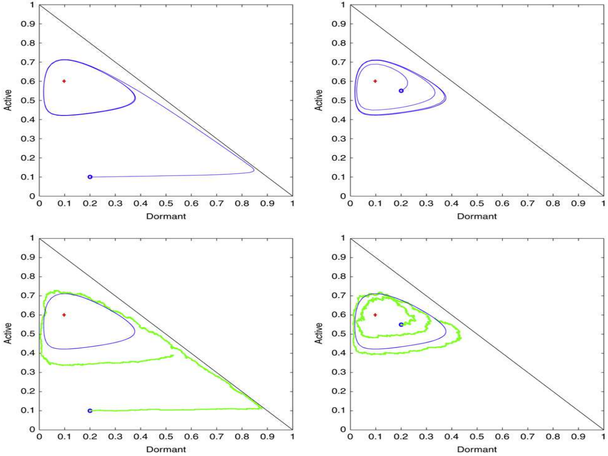
$$\Lambda(\xi)^T \xi = 0.$$

- If the solution is unique, say ξ^*, predict that the system will settle down at $\xi^* \otimes \xi^* \otimes \ldots \otimes \xi^*$.
- Works very well for the exponential backoff.
- Another example in the next slide

SIS system and “herd immunity”

- Normalise time so that recovery rate is 1. Assume that the contact rate is β.
- In this normalisation, $\beta = R_0$ of the infection.
- The model is $\dot{\mu}_1(t) = \beta\mu_1(t)(1 - \mu_1(t)) - \mu_1(t)$, with $\mu(0) = \nu$.
- Rest points ξ^* solve $\beta\xi^*(1 - \xi^*) - \xi^* = 0$
- $\xi^* = 0$ or $\xi^* = 1 - 1/\beta$ (not every one in infected state).

Issues: A malware propagation example from Benaim and Le Boudec 2008



- The fixed point is unique, but unstable.
- All trajectories starting from outside the fixed point, and all trajectories in the finite N system, converge to the stable limit cycle.

A sufficient condition when the method works

Theorem

Assume that the transition graph forms one communicating class and assume Lipschitz rates.

Let $\mu_N(0) \to \nu$ in probability.

Let the ODE have a (unique) globally asymptotically stable equilibrium ξ^* with every path tending to ξ^*.

Then $\mu_N(\infty) \xrightarrow{d} \xi^*$.

It is not enough to have a unique fixed point ξ^*.
But if that ξ^* is globally asymptotically stable, that suffices.

A sufficient condition

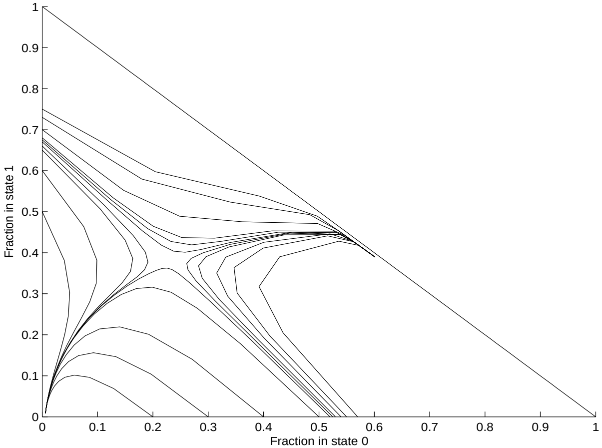
A lot of effort has gone into identifying when we can ensure a globally asymptotic stable equilibrium.

Theorem

If c is such that $\langle \xi, c \rangle < 1$ for all ξ, then the rest point ξ^* of the dynamics is unique, and all trajectories converge to it.

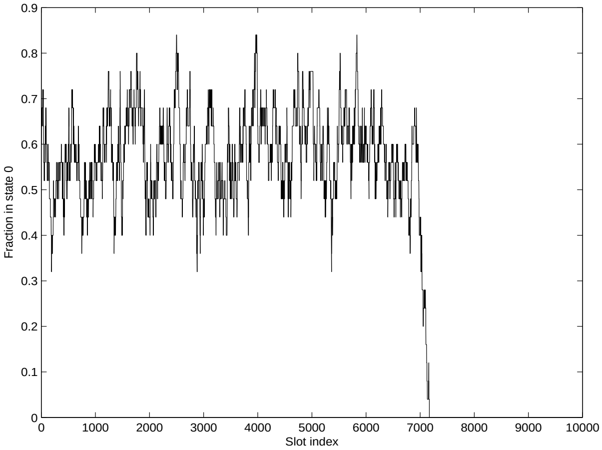
This is the case for the classical exponential backoff with $c_0 < 1$.

The case of multiple stable equilibria for the ODE



- Different parameters: $c = (0.5, 0.3, 8.0)$.
- There are two stable equilibria.
  One near $(0.6, 0.4, 0.0)$ and another near $(0, 0, 1)$.

The case of multiple stable equilibria: metastability



Fraction of nodes in state 0 is near 0.6 for a long time, but then moves to 0, and in a sequence of rapid steps.

The reverse move is a lot less frequent.

A selection principle: Preview to the second hour

- If unique globally asymptotically stable equilibrium ξ^*, then $\mu_N(\infty) \xrightarrow{d} \xi^*$. (Limit law).

- If we encounter multiple stable limit sets, look at probability of a large deviation.

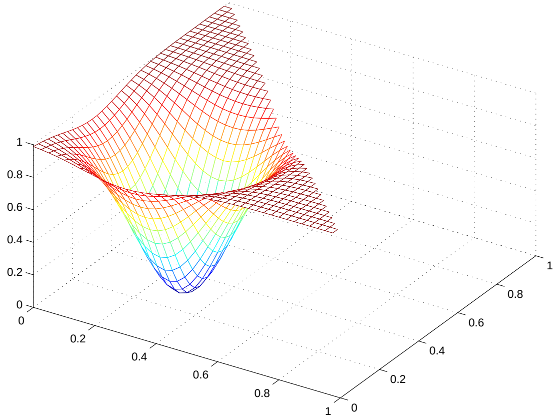
- Characterise the exponent in

$$\Pr\{\mu_N(\infty) \in \text{ neighbourhood of } \xi\} \sim \exp\{-NV(\xi)\}.$$

- The locations $\{\xi : V(\xi) = 0\}$ should "select" the correct limit set.

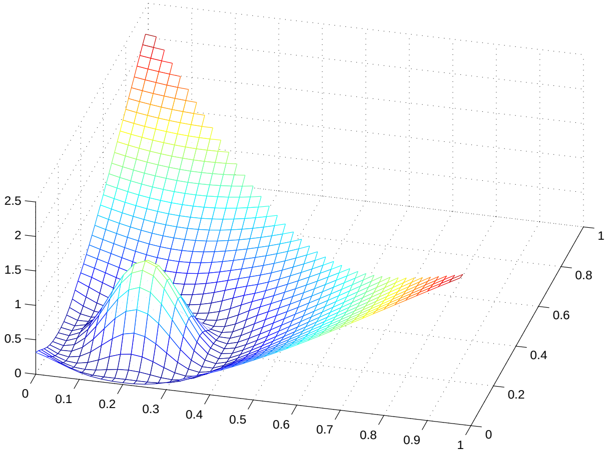
- $V(\xi)$ is called a quasipotential (Freidlin-Wentzell).

Quasipotential $V(\xi)$



The case of a (unique) globally asymptotically stable equilibrium for the McKean-Vlasov dynamics: $V(\xi^*) = 0$.

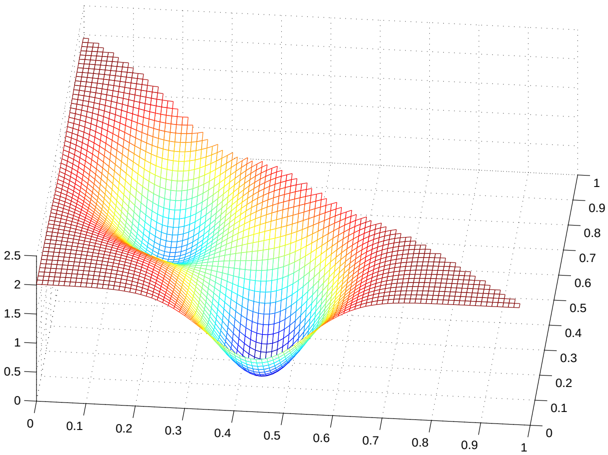
Quasipotential $V(\xi)$



The case of a unique but unstable rest point. $V(\xi^*) > 0$.

All trajectories converge to the stable limit cycle.

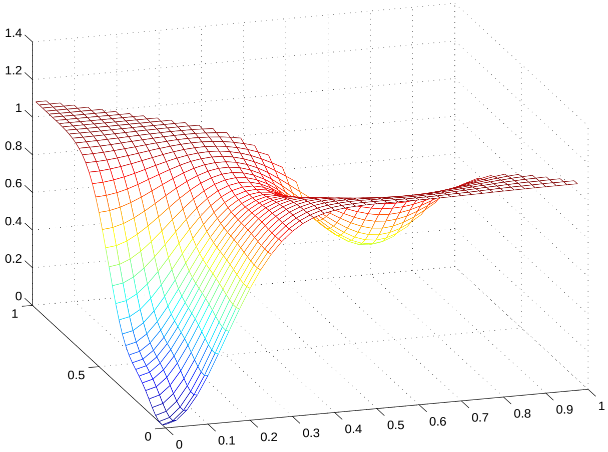
Quasipotential $V(\xi)$



The case of two stable equilibria.

The selection is the one that has the deepest shade of blue ($V(\xi_1^*) = 0$).

Quasipotential $V(\xi)$



A qualitative picture for the case $c = (0.5, 0.3, 8.0)$.

The two stable points are $(0.6, 0.4, 0.0)$ and $(0.0, 0.0, 1.0)$.
The latter is a truer representative of the large time behaviour.

Proofs: First Kurtz's theorem

Theorem

Suppose that the initial empirical measure $\mu_N(0) \xrightarrow{p} \nu$, where ν is deterministic.

Assume each $\lambda_{i,j}(\cdot)$ is Lipschitz in its argument. Let $\mu(\cdot)$ be the solution to the McKean-Vlasov dynamics with initial condition $\mu(0) = \nu$.

Then $\mu_N(\cdot) \xrightarrow{p} \mu(\cdot)$.

Technicalities:

- Fix $T > 0$ and $\varepsilon > 0$. We will argue

$$\begin{aligned} \Pr\{d_T(\mu_N, \mu) > \varepsilon\} &\leq \Pr\{\|\mu_N(0) - \mu(0)\| > \varepsilon/(2e^{MT})\} \\ &\quad + C_1 \exp\{-NT\overline{\lambda} h(\varepsilon/(C_2 T e^{MT}))\} \end{aligned}$$

  where M is the Lipschitz constant of the driving function, $\overline{\lambda}$ is the max of the transition rates, and
  $h(t) = (1+t)\ln(1+t) - t, \quad t > -1.$

Proofs: Proof of Kurtz's theorem

- Time change. Let $M(\cdot)$ be a unit rate Poisson point process (PPP). Then $M(\int_0^\cdot \lambda(s)ds)$ is a time-inhomogeneous PPP with instantaneous rate $\lambda(\cdot)$.
- Let $(M_{i,j}(\cdot))_{i,j}$ be independent unit-rate PPP.

$$
\begin{aligned}
\mu_N(t) &= \mu_N(0) + \sum_{i,j} \left(\frac{\delta_j - \delta_i}{N}\right) M_{i,j}\left(\int_0^t N\mu_N(s)(i)\lambda_{i,j}(\mu_N(s))ds\right) \\
&= \mu_N(0) + \int_0^t F(\mu_N(s))ds + \sum_{i,j} \left(\frac{\delta_j - \delta_i}{N}\right) \overline{M}_{i,j}(\cdot)
\end{aligned}
$$

- Martingale noise $\overline{M}_{i,j}(t)$ is of the form $M_{i,j}(t) - t$
- By triangle inequality and Lipschitz,

$$
\begin{aligned}
\|\mu_N(t) - \mu(t)\| &\leq \|\mu_N(0) - \mu(0)\| + \int_0^t \|F(\mu_N(s)) - F(\mu(s))\| ds + \|noise\| \\
&\leq \|\mu_N(0) - \mu(0)\| + M \int_0^t \|\mu_N(s)) - \mu(s)\| ds + \|noise\|
\end{aligned}
$$

- Then Poisson concentration and Gronwall.

Proofs: Marginal

$X_{n_1}^{(N)}(t) \xrightarrow{d} U_1$ where U_1 is a random variable with distribution $\mu(t)$.

- Take any bounded test function ϕ on $\mathcal{Z}$.
- Suffices to show $\mathbb{E}[\phi(X_{n_1}^{(N)}(t))] \to \mathbb{E}[\phi(U_1)]$

$$\begin{aligned}
\mathbb{E}[\phi(X_{n_1}^{(N)}(t))] &= \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N}\phi(X_n^{(N)}(t))\right] \\
&= \mathbb{E}\left[\langle \mu_N(t), \phi\rangle\right] \\
&\to \langle \mu(t), \phi\rangle \\
&= \mathbb{E}[\phi(U_1)]
\end{aligned}$$

Proofs: Double marginal

$(X_{n_1}^{(N)}(t), X_{n_2}^{(N)}(t)) \xrightarrow{d} (U_1, U_2)$, where U_1 and U_2 are iid $\sim \mu(t)$.

- Take any two bounded test functions ϕ_1 and ϕ_2 on $\mathcal{Z}$.
- Suffices to show $\mathbb{E}[\phi_1(X_{n_1}^{(N)}(t))\phi_2(X_{n_1}^{(N)}(t)] \to \mathbb{E}[\phi_1(U_1)]\ \mathbb{E}[\phi_2(U_2)]$

$$
\begin{aligned}
&\mathbb{E}[\phi_1(X_{n_1}^{(N)}(t))\phi_2(X_{n_1}^{(N)}(t))] - \mathbb{E}[\phi_1(U_1)]\ \mathbb{E}[\phi_2(U_2)] \\
= \;& \mathbb{E}\left[\phi_1(X_{n_1}^{(N)}(t))\phi_2(X_{n_1}^{(N)}(t))\right] - \mathbb{E}\left[\prod_{l=1}^{2}\langle \mu_N(t), \phi_l\rangle\right] \\
& + \mathbb{E}\left[\prod_{l=1}^{2}\langle \mu_N(t), \phi_l\rangle\right] - \mathbb{E}\left[\phi_1(U_1)\right]\ \mathbb{E}\left[\phi_2(U_2)\right] \\
= \;& \mathbb{E}\left[\frac{1}{N(N-1)}\sum_{n_1 \neq n_2}\phi_1(X_{n_1}^{(N)}(t))\phi_2(X_{n_1}^{(N)}(t))\right] \\
& - \mathbb{E}\left[\left(\frac{1}{N}\sum_{n_1}\phi_1(X_{n_1}^{(N)}(t))\right)\left(\frac{1}{N}\sum_{n_2}\phi_2(X_{n_2}^{(N)}(t))\right)\right] \\
& + \mathbb{E}\left[\prod_{l=1}^{2}\langle \mu_N(t), \phi_l\rangle\right] - \prod_{l=1}^{2}\langle \mu(t), \phi_l\rangle
\end{aligned}
$$

Proofs: Globally asymptotically stable equilibrium and stationary regime

Globally asymptotically stable equilibrium $\Rightarrow \mu_N(\infty) \xrightarrow{d} \xi^*$.

- $\pi_N := Law(\mu_N(0))$, invariant measure. Then $\pi_N = Law(\mu_N(t))$ also.
- Compactness implies subsequential limits $\pi_{N_l} \to \pi$.
- $\pi = \pi \circ \Phi_t^{-1}$, under the McKean-Vlasov flow Φ_t
- Compactness of the space, Liapunov stability, Gronwall implies that for every $\varepsilon > 0$, there is a T such that $\forall t > T$, we have support of $(\pi \circ \Phi_t^{-1}) \subset B_\varepsilon(\xi^*)$ for all $t > T$.
- So support of π is within a ball of ε around ξ^*.
- $\varepsilon > 0$ is arbitrary. So support of π is $\{\xi^*\}$ and $\pi = \delta_{\xi^*}$, unique.

Section 2

Large deviations of mean-field models

Mean-Field Interacting Particle Systems: Limit Laws and Large Deviations

Section 2: Large Deviations of Mean-Field Models

SIGMETRICS/PERFORMANCE 2022

Recall the mean-field model

- N particles. The state of the nth particle is $X_n^N(t) \in \mathcal{Z}$. The empirical measure at time t is

$$\mu_N(t) = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(t)}.$$

- An $i \to j$ transition occurs at rate $\lambda_{i,j}(\mu_N(t))$.
- The McKean-Vlasov equation:

$$\dot{\mu}_t = \Lambda(\mu_t)^T \mu_t, \; t \geq 0.$$

- We will now quantify various rare events associated with $\{\mu_N\}$.

Outline of Section 2

- An introduction to large deviations.
  - Basic definitions, some examples.
- Process-level large deviations of the family $\{\mu_N\}$.
  - A change of measure argument.
- Large deviations of the invariant measure of μ_N.

A primer on large deviations

Large deviation principle (LDP)

- Let S be a complete and separable metric space. Let $\{X_N, N \geq 1\}$ be a sequence of S-valued random variables.
- Roughly, $P(X_N \in A) \sim \exp\{-N \inf_{x \in A} I(x)\}$.
- Here, $I : S \to [0, \infty]$ is called the rate function.

Large deviation principle (LDP)

Definition

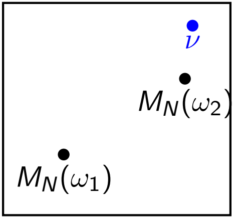
$\{X_N, N \geq 1\}$ is said to satisfy the LDP on S with rate function I if

- (Compactness of level sets). For any $s \geq 0$, $\Phi(s) := \{x \in S : I(x) \leq s\}$ is a compact subset of S;
- (LDP lower bound). For any $\gamma > 0$, $\delta > 0$, and $x \in S$, there exists $N_0 \geq 1$ such that
$$P(d(X_N, x) < \delta) \geq \exp\{-N(I(x) + \gamma)\}$$
for any $N \geq N_0$;
- (LDP upper bound). For any $\gamma > 0$, $\delta > 0$, and $s > 0$, there exists $N_0 \geq 1$ such that
$$P(d(X_N, \Phi(s)) \geq \delta) \leq \exp\{-N(s - \gamma)\}$$
for any $N \geq N_0$.

Example: Sanov's theorem

- Let S be a Polish space. Let μ be a probability measure on S.
- Let $X_1, X_2, \ldots, X_N$ be i.i.d. μ.
- Define the empirical measure

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N} \delta_{X_n}.$$



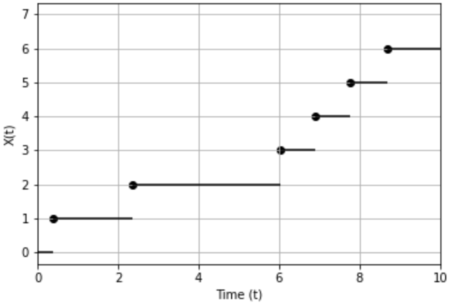
- This is an $\mathcal{M}_1(S)$-valued random variable.
- By the weak law of large numbers, $\mu_N \to \mu$ in $\mathcal{M}_1(S)$ as $N \to \infty$, in probability.
- But there is a positive probability for μ_N to be close to $\nu \neq \mu$.

Theorem (Sanov)

$\{\mu_N, N \geq 1\}$ *satisfies the LDP on* $\mathcal{M}_1(S)$ *with rate function* $I(\cdot\|\mu)$.

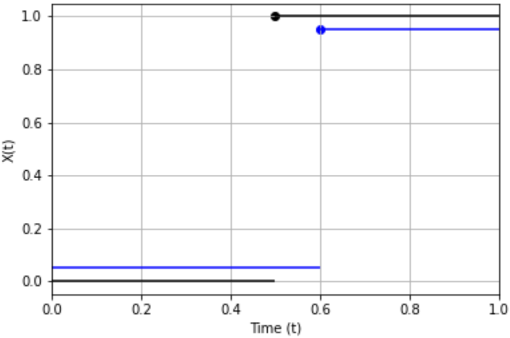
The D-space

- Let S be a complete and seperable metric space.
- Fix $T > 0$. Let $D([0, T], S)$ denote the space of S-valued functions on $[0, T]$ that are
  - Right continuous at each $t \in [0, T)$, and
  - Possesses left limits at each $t \in (0, T]$.
- Examples:
  - All continuous functions on $[0, T]$.
  - Trajectories of a Poisson point process.



The D-space

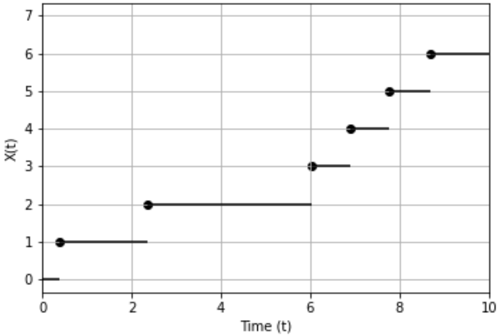
- We can define a distance function on D that takes into account small time perturbations.



- Under this metric, D is a complete and seperable metric space.

Example: LDP on the space of trajectories

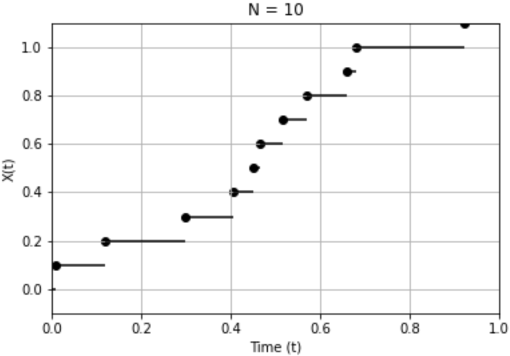
- Consider the unit rate Poisson point process $X(t)$ for $t \in [0, T]$.



- X is a $D([0, T], \mathbb{R})$-valued random variable.

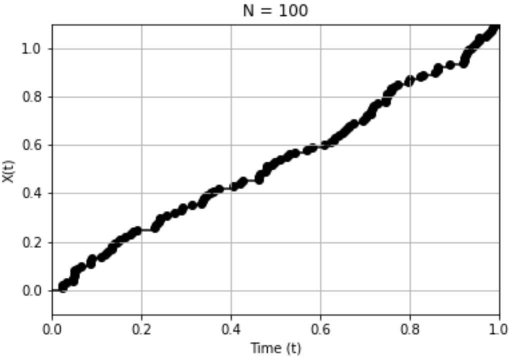
Example: LDP on the space of trajectories

- Consider the time-scaled and amplitude-scaled process: $\frac{1}{N}X(Nt)$.



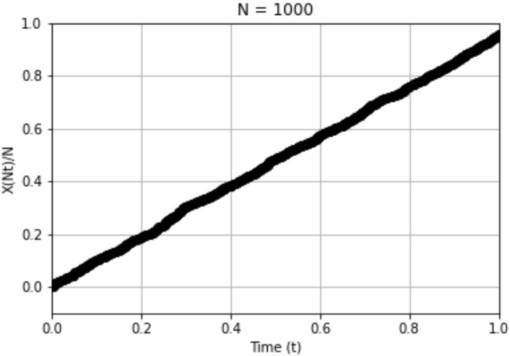
Example: LDP on the space of trajectories

- Consider the time-scaled and amplitude-scaled process: $\frac{1}{N}X(Nt)$.



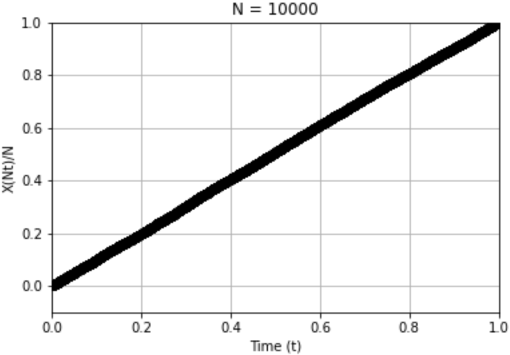
Example: LDP on the space of trajectories

- Consider the time-scaled and amplitude-scaled process: $\frac{1}{N} X(Nt)$.



Example: LDP on the space of trajectories

- Consider the time-scaled and amplitude-scaled process: $\frac{1}{N}X(Nt)$.

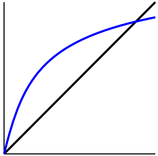


- The process $\frac{1}{N}X(Nt)$ is a small random perturbation of the ODE

$$\dot{x}(t) = 1, \; x(0) = 0, \; t \in [0, 1].$$

Example: LDP on the space of trajectories

- Question: probability that $\frac{1}{N}X(Nt)$ tracks a given function φ?



- One can show that $\{\frac{1}{N}X(Nt), N \geq 1\}$ satisfies the LDP on $D([0,T],\mathbb{R})$ with rate function

$$S(\varphi) = \int_{[0,T]} \tau^*(\dot{\varphi}(t) - 1)dt,$$

if $t \mapsto \varphi(t)$ is absolutely continuous, increasing, and $\varphi(0) = 0$; $S(\varphi) = \infty$ otherwise.

- Here,

$$\tau^*(x) = \begin{cases} (x+1)\log(x+1) - x, & \text{if } x \geq -1, \\ \infty, & \text{if } x < -1. \end{cases}$$

A closer look at the rate function

- $$S(\varphi) = \int_{[0,T]} \tau^*(\dot{\varphi}(t) - 1)dt.$$
- τ^* is the convex dual of $\tau(u) = e^u - u - 1$, $u \in \mathbb{R}$;
$$\tau^*(t) = \sup_u (ut - \tau(u)), \ t \in \mathbb{R}.$$
- So,
$$S(\varphi) = \int_{[0,T]} \sup_u (u(\dot{\varphi}(t) - 1) - \tau(u))dt.$$
- Such variational forms will appear later.

Contraction principle

- S, T are metric spaces. $f : S \to T$ is continuous.
- $\{X_N\}$s are S-valued random variables. Define $Y_N = f(X_N)$.

Theorem (Contraction Principle)

If $\{X_N\}$ satisfies the LDP with rate function I, then $\{Y_N\}$ satisfies the LDP with rate function

$$J(y) = \inf_{x \in S : y = f(x)} I(x).$$

A new LDP from change of measure

- Let $\{P_N\}$ satisy the LDP with rate function I.
- Let Q_N be such that

$$\frac{dQ_N}{dP_N}(x) = \exp\{Nf(x)\},$$

  for some $f : S \to \mathbb{R}$, bounded and continuous.
- Additionally, suppose that $\{Q_N\}$ is exponentially tight: Given $M > 0$, there exists a compact set K_M such that $Q_N(K_M^c) \leq \exp\{-NM\}$ for all N.
- Then, $\{Q_N\}$ satisfies the LDP with rate function $I(x) - f(x)$.

A new LDP from change of measure

- Lower bound: For $x \in S$ and $\delta > 0$,

$$\begin{aligned} Q_N(d(X_N, x) < \delta) &= E^{Q_N}(\mathbf{1}_{\{X_N \in B(x,\delta)\}}) \\ &= E^{P_N}(\exp\{Nf(X_N)\}\mathbf{1}_{\{X_N \in B(x,\delta)\}}) \\ &\geq \exp\{N(f(x) - \varepsilon)\}P_N(X_N \in B(x,\delta)) \\ &\geq \exp\{-N(I(x) - f(x) + 2\varepsilon)\}. \end{aligned}$$

- Upper bound: For a closet set F,

$$\begin{aligned} Q_N(F) &\leq Q_N(K_M^c) + Q_N(F \cap K_M) \\ &\leq \exp\{-NM\} + Q_N(F \cap K_M). \end{aligned}$$

- Since $F \cap K_M$ is compact, we can cover it using a finite number of balls. For the ith ball,

$$Q_N(\overline{B}(x_i, \delta)) \leq \exp\{-N(I(x) - f(x) - \varepsilon)\}.$$

Varadhan's lemma

Theorem

Let $f : S \to \mathbb{R}$ be bounded and continuous. Suppose that $\{X_N\}$ satisfies the LDP with rate function I. Then

$$\lim_{N\to\infty} \frac{1}{N} \log E(\exp\{Nf(X_N)\}) = \sup_{x\in S}(f(x) - I(x)).$$

- By the LDP, $E(\exp\{Nf(X_N)\}\mathbf{1}_{\{X_N\sim x\}}) \sim \exp\{Nf(x)\}\exp\{-NI(x)\}$.
- The leading terms in the expectation are those $x \in S$ for which $f(x) - I(x)$ is the largest.

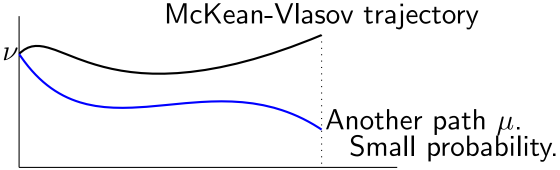
Large deviations of the empirical measure process

Recall the empirical measure process

- $\mu_N(t) \to \mu_N(t) + \frac{\delta_j}{N} - \frac{\delta_i}{N}$ at rate $N\mu_N(t)(i)\lambda_{i,j}(\mu_N(t))$.
- Recall the McKean-Vlasov equation:

$$\dot{\mu}_t = \Lambda(\mu_t)^T \mu_t, \; t \geq 0.$$

- From Section 1, if $\mu_N(0) \to \nu$ in $\mathcal{M}_1(\mathcal{Z})$, then $\mu_N(\cdot) \to \mu(\cdot)$ in $D([0,T], \mathcal{M}_1(\mathcal{Z}))$, in probability.
- We now present the large deviations of μ_N.



Large deviations of μ_N

Theorem

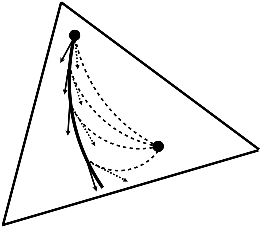
Let $\mu_N(0) \to \nu$ in $\mathcal{M}_1(\mathcal{Z})$. Then μ_N satisfies the LDP on $D([0,T], \mathcal{M}_1(\mathcal{Z}))$ with rate function $S_{[0,T]}(\cdot|\nu)$ defined as follows. If $\mu_0 = \nu$ and $[0,T] \ni t \mapsto \mu_t \in \mathcal{M}_1(\mathcal{Z})$ is absolutely continuous,

$$S_{[0,T]}(\mu|\nu) = \int_{[0,T]} \sup_{\alpha \in \mathbb{R}^{|\mathcal{Z}|}} \Bigg\{ \langle \alpha, \dot{\mu}_t - \Lambda(\mu_t)^T \mu_t \rangle - \sum_{(i,j) \in \mathcal{E}} \tau(\alpha(j) - \alpha(i)) \lambda_{i,j}(\mu_t) \mu_t(i) \Bigg\} dt,$$

else $S_{[0,T]}(\mu|\nu) = \infty$. Here, $\tau(u) = e^u - u - 1$.

An interpretation of the rate function

- Consider a path $\dot{\mu}_t = G(t)^T \mu_t$.



- In a small time around t, for an $i \to j$ transition,
  - The usual rate is Bernoulli($p = \lambda_{i,j}(\mu(t))dt$).
  - The new rate is Bernoulli($q = G_{i,j}(t)dt$).
- By Sanov's theorem, the infinitesimal cost of this change is

$$I(\text{Bernoulli}(q) \| \text{Bernoulli}(p)) = \left(q \log \frac{q}{p} - q + p \right).$$

- Accumulate these costs over $[0, T]$ to get the rate function.

LDP for $\{\mu_N\}$ – proof sketch

- Consider a system of non-interacting particles.
  - $\lambda_{i,j}(\xi) = 1$ for all $\xi \in \mathcal{M}_1(\mathcal{Z})$ and $(i,j) \in \mathcal{E}$.
- Define the empirical measure on paths

$$\overline{\mu}_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N}.$$

- This is a $\mathcal{M}_1(D([0,T],\mathcal{Z}))$ valued random variable.
  - $\overline{\mu}_N(t) = \bar{\mu} \circ \pi_t^{-1}$, where π_t is the projection mapping

$$D([0,T],\mathcal{Z}) \ni \varphi \mapsto \varphi(t) \in \mathcal{M}_1(\mathcal{Z}).$$

- Let $\bar{P}_z$ denote the law of a particle starting at z.
- If $X_n^N(0) = z$ for all n, then by Sanov's theorem, $\{\overline{\mu}_N\}$ satisfies the LDP with rate function $Q \mapsto I(Q \| \overline{P}_z)$.

LDP for $\{\mu_N\}$ – proof sketch

- When $\overline{\mu}_N(0) \to \nu$, then a generalisation of Sanov's theorem gives the LDP for $\{\overline{\mu}_N\}$ with rate function

$$J(Q) = \sup_{f \in C_b(D)} \left[\int_D f dQ - \sum_{z \in \mathcal{Z}} \nu(z) \log \int_D e^f d\overline{P}_z \right]$$

  (Dawson and Gärtner, 1987).
- In particular, when $\nu = \delta_z$, $J(Q) = I(Q \| \bar{P}_z)$.
- By Jensen's inequality, $J(Q) \geq I(Q \| \sum_z \nu(z) \bar{P}_z)$.

A change of measure

- Consider two probability measures: $P \sim \text{Poisson}(\lambda_1)$, and $Q \sim \text{Poisson}(\lambda_2)$.
- We have

$$P(k) = \frac{\lambda_1^k \exp\{-\lambda_1\}}{k!}, \ k \geq 0,$$

and similarly $Q(k)$.

- So,

$$\begin{aligned} \frac{Q(k)}{P(k)} &= \left(\frac{\lambda_2}{\lambda_1}\right)^k \exp\{-(\lambda_2 - \lambda_1)\} \\ &= \exp\left\{k \log\left(\frac{\lambda_2}{\lambda_2}\right) - (\lambda_2 - \lambda_1)\right\}. \end{aligned}$$

A change of measure

- More generally, let P (resp. Q) be the law of the Poisson point process with rate λ_1 (resp. λ_2).
- Both P and Q are probability measures on $D([0,T],\mathbb{Z}_+)$.
- By Girsanov's theorem,

$$\frac{dQ}{dP}(x) = \exp\left\{ \sum_{0\le t\le T} \mathbf{1}_{\{x_t \neq x_{t-}\}} \log\left(\frac{\lambda_2}{\lambda_1}\right) - \int_{[0,T]} (\lambda_2 - \lambda_1) dt \right\},$$

for $x \in D([0,T],\mathbb{Z}_+)$.

LDP for $\{\mu_N\}$ – proof sketch

- Let $\mathbb{P}_N$ (resp. $\overline{\mathbb{P}}_N$) be the law of the interacting (resp. non-interacting) system.
- By Girsanov's theorem,

$$\frac{d\mathbb{P}_N}{d\overline{\mathbb{P}}_N}(Q) = \exp\{Nh(Q)\}, Q \in \mathcal{M}_1(D),$$

where,

$$h(Q) = \int_D h_1(x, Q) Q(dx),$$

$$h_1(x, Q) = \sum_{0 \le t \le T} \mathbf{1}_{\{x_t \neq x_{t-}\}} \log \lambda_{x_{t-}, x_t}(Q(t-)) - \int \sum_{j:(x_{t-}, j) \in \mathcal{E}} (\lambda_{x_{t-}, j}(Q(t-)) - 1) dt.$$

LDP for $\{\mu_N\}$ – proof sketch

- However, h is neither bounded nor continuous.
- Consider a subspace of $\mathcal{M}_1(D)$:

$$M_{1,\varphi}(D) = \left\{ Q \in \mathcal{M}_1(D) : \int_D \varphi dQ < \infty \right\},$$

  where, $\varphi : D \to \mathbb{R}_+$ is the function $\varphi(x) = \sum_{0 \le t \le T} \mathbf{1}_{\{x_t \neq x_{t-}\}}$.
- Show that h is continuous at all points in $M_{1,\varphi}(D)$.
- Then show that $\{\gamma_N\}$ satisfies the LDP with rate function $Q \mapsto J(Q) - h(Q)$.
- By the contraction principle, $\{\mu_N(t)\}$ satisfies the LDP with rate function $S_{[0,T]}(\cdot|\nu)$.

Large deviations in the stationary regime

The unique attractor case

- Recall the empirical measure process μ_N. Let $\wp_N$ be its unique invariant probability measure.
- $\wp_N$ is the law of $\mu_N(\infty)$. It is a probability measure on $\mathcal{M}_1(\mathcal{Z})$.
- Recall the McKean-Vlasov equation

$$\dot{\mu}_t = \Lambda(\mu_t)^T \mu_t, \ t \geq 0.$$

- Suppose that ξ^* is the unique globally asymptotically stable equilibrium of the McKean-Vlasov equation.
- From Section 1, $\mu_N(\infty)$ converges to ξ^* in distribution as $N \to \infty$.
- We now study the large deviations of $\{\wp_N\}$.

LDP for the terminal time

- Consider the random variable $\mu_N(T)$.
- The mapping

$$D([0,T],\mathcal{M}_1(\mathcal{Z})) \ni \varphi \mapsto \varphi(T) \in \mathcal{M}_1(\mathcal{Z})$$

  is continuous.
- Let $\mu_N(0) \to \nu$. By the contraction principle, $\{\mu_N(T)\}$ satisfies the LDP with rate function

$$S_T(\xi|\nu) = \inf\{S_{[0,T]}(\mu|\nu) : \mu(0) = \nu, \mu(T) = \xi\}.$$

LDP for the joint law $(\mu_N(0), \mu_N(T))$

- So far, we assumed $\mu_N(0) \to \nu$.
- Suppose we start at stationarity, i.e., the law of $\mu_N(0)$ is $\wp_N$. Then the law of $\mu_N(T)$ is also $\wp_N$.
- Consider $(\mu_N(0), \mu_N(T))$.
- Suppose that $\wp_N$ satisfies the LDP with rate function V. Then, under some conditions, the joint law $(\mu_N(0), \mu_N(T))$ satisfies the LDP with rate function

$$(\nu, \xi) \mapsto V(\nu) + S_T(\xi|\nu)$$

A recursion for the rate function

- Suppose that $\wp_N$ satisfies the LDP with rate function V.
- We have that $(\mu_N(0), \mu_N(T))$ satisfies the LDP with rate function

$$(\nu, \xi) \mapsto V(\nu) + S_T(\xi|\nu)$$

- On one hand, by the contraction principle, $\{\mu_N(T)\}$ satisfies the LDP with rate function

$$\xi \mapsto \inf_{\nu \in \mathcal{M}_1(\mathcal{Z})} [V(\nu) + S_T(\xi|\nu)]$$

- On the other hand, since the law of $\mu_N(T)$ is $\wp_N$, we have

$$V(\xi) = \inf_{\nu \in \mathcal{M}_1(\mathcal{Z})} [V(\nu) + S_T(\xi|\nu)] \text{ for all } T > 0.$$

- Is there a unique V that satisfies this?

Large deviations of $\wp_N$

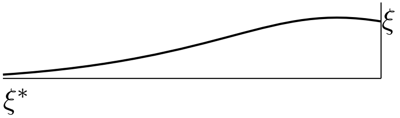
Theorem

The family $\{\wp_N\}$ satisfies the LDP on $\mathcal{M}_1(\mathcal{Z})$ with rate function

$$V(\xi) = \inf_{T>0} S_T(\xi|\xi^*).$$

Further, there exists a trajectory $\hat{\mu}$ such that $\hat{\mu}(t) \to \xi^*$ as $t \to -\infty$, $\hat{\mu}(0) = \xi$, and

$$V(\xi) = S_{(-\infty,0]}(\hat{\mu}|\xi^*).$$



Large deviations of $\wp_N$ – proof sketch

- Show that $V(\xi^*) = 0$.
- Then,

$$V(\xi) \leq V(\xi^*) + S_T(\xi|\xi^*) \text{ for all } T > 0.$$

- So,

$$V(\xi) \leq \inf_{T>0} S_T(\xi|\xi^*).$$

Large deviations of $\wp_N$ – proof sketch

- For $T > 0$, show that the infimum in

$$\inf_{\nu \in \mathcal{M}_1(\mathcal{Z})} [V(\nu) + S_T(\xi|\nu)]$$

  is attained.
- For each ν, ξ and $T > 0$, there is an optimal path $\hat{\mu}$ from ν to ξ, i.e., $S_T(\xi|\nu) = S_{[0,T]}(\hat{\mu}|\nu)$.
- So,

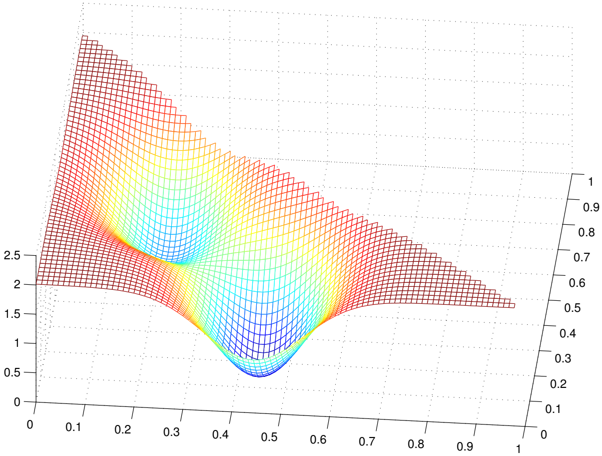
$$V(\xi) = V(\hat{\mu}(-mT)) + S_{mT}(\xi|\hat{\mu}(-mT)).$$

- Argue that $\hat{\mu}(-mT) \to \xi^*$ as $m \to \infty$.
- By the lower semicontinuity of V, and $V(\xi^*) = 0$, we have

$$V(\xi) \geq S_{(-\infty,0]}(\hat{\mu}|\xi^*).$$

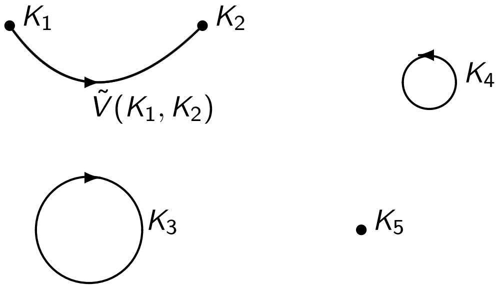
The general case – multiple equilibria

- The Freidlin-Wentzell quasipotential V on $\mathcal{M}_1(\mathcal{Z})$.
- $P(\mu_N(\infty) \sim \xi) \sim \exp\{-NV(\xi)\}$.



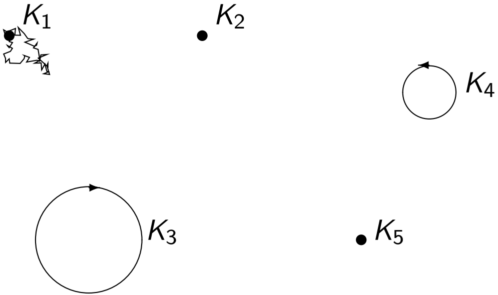
The general case – some notation

- Assumptions on the McKean-Vlasov equation: There exists a finite number of compact sets $K_1, K_2, \ldots, K_l$ such that
  - Every equilibrium of the McKean-Vlasov equation lies completely in one of the compact sets K_i.
  - No cost of movement within K_i. Positive cost to go out of (or come into) K_i.

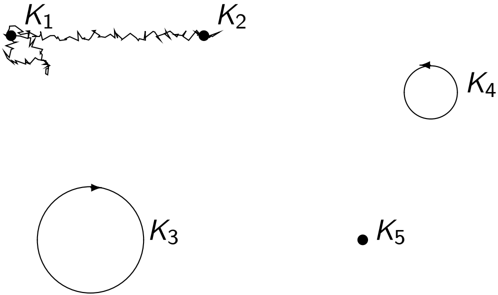


- $\tilde{V}(K_i, K_j) = \inf\{S_{[0,T]}(\varphi|\varphi_0) : \varphi_0 \in K_i, \varphi_T \in K_j, \varphi_t \notin \cup_{i' \neq i,j} K_{i'}, T > 0\}$ (communication cost from K_i to K_j).

Approximation of μ_N using a discrete chain



Approximation of μ_N using a discrete chain



- τ_n: hitting time of μ_N in a given neighbourhood of K_i's.
- Hitting time chain: $Z_n^N = \mu_N(\tau_n)$, $n \geq 1$.
- To quantify the transitions between K_i's, we need large deviation estimates of μ_N *uniformly* with respect to the initial condition.

Uniform large deviations

- μ_N^ν: process starting from ν. Indexed by two parameters.

Definition

$\{\mu_N^\nu\}$ is said to satisfy the uniform LDP over a class of subsets $\mathcal{A} \subset \mathcal{M}_1(\mathcal{Z})$ if

- for each $K \subset \mathcal{M}_1(\mathcal{Z})$ compact and $s > 0$, $\mathcal{K} = \bigcup_{\nu \in K} \Phi_\nu(s)$ is a compact subset of $D([0,T], \mathcal{M}_1(\mathcal{Z}))$;
- for any $\gamma > 0, \delta > 0, s > 0$ and $A \in \mathcal{A}$, there exists $N_0 \geq 1$ such that

  $$P_\nu(\rho(\mu_N^\nu, \varphi) < \delta) \geq \exp\{-N(S_{[0,T]}(\varphi|\nu) + \gamma)\},$$

  for all $\nu \in A$, $\varphi \in \Phi_\nu(s)$ and $N \geq N_0$;
- for any $\gamma > 0, \delta > 0, s_0 > 0$ and $A \in \mathcal{A}$, there exists $N_0 \geq 1$ such that

  $$P_\nu(\rho(\mu_N^\nu, \Phi_\nu(s)) \geq \delta) \leq \exp\{-N(s - \gamma)\},$$

  for all $\nu \in A$, $s \leq s_0$ and $N \geq N_0$.

- Theorem: $\{\mu_N^\nu\}$ satisfies the uniform LDP over $\mathcal{M}_1(\mathcal{Z})$.

One step transition probability of Z^N

Lemma

Given $\varepsilon > 0$, there exists $\delta > 0$ such that the one-step transition probability of the chain Z^N satisfies

$$\exp\{-N(\tilde{V}(K_i, K_j) + \varepsilon)\} \leq P(B(K_i, \delta), B(K_j, \delta)) \leq \exp\{-N(\tilde{V}(K_i, K_j) - \varepsilon)\}$$

for all large enough N.

- Upon exit from K_i, μ_N is most likely to visit K_j that attains $\min_{j'} \tilde{V}(K_i, K_{j'})$ $(= \tilde{V}(K_i))$.

One step transition probability of Z^N – proof sketch

- Lower bound:
  - By the definition of $\tilde{V}(K_i, K_j)$, given $\varepsilon > 0$, there exists a trajectory φ from K_i to K_j such that $S_{[0,T]}(\varphi|K_i) \leq \tilde{V}(K_i, K_j) + \varepsilon$.
  - Then, using the uniform LDP for $\{\mu_N\}$,

$$\begin{aligned} P(B(K_i,\delta), B(K_j,\delta)) &\geq P_{K_i}(\mu_N \in \text{nbhd}(\varphi)) \\ &\geq \exp\{-N(\tilde{V}(K_i,K_j)+\varepsilon)\}. \end{aligned}$$

- Upper bound:
  - Let τ_1 be the hitting time of $\cup K_l$.
  - Given $M > 0$, we can find $T_1 > 0$ such that $P_{K_i}(\tau_1 > T_1) \leq \exp\{-NM\}$.
  - Let $A = \{\varphi : \varphi_0 \in K_i, \varphi_{T_1} \in K_j, S_{[0,T]}(\varphi|K_i) \leq \tilde{V}(K_i, K_j) - \varepsilon\}$.
  - Then using the uniform LDP for $\{\mu_N\}$,

$$\begin{aligned} P(B(K_i,\delta), B(K_j,\delta)) &\leq P_{K_i}(\tau_1 \geq T_1) + P_{K_i}(\text{dist}(\mu_N, A) \geq \delta) \\ &\leq \exp\{-NM\} + \exp\{-N(\tilde{V}(K_i,K_j)-\varepsilon)\}. \end{aligned}$$

The Markov chain tree theorem

- Consider an irreducible Markov chain on $L = \{1, 2, \ldots, l\}$ with transition probaiblity matrix P.
- An i-graph $G(i)$ is a directed graph on L such that
  - There is exactly one outgoing arrow from every $j \in L$.
  - There are no closed cycles.
- For an i-graph g, let $\pi(g) = \prod_{(i,j)\in g} P(i,j)$.
- Let $W(i) = \sum_{g \in G(i)} \pi(g)$.
- Then,

$$\frac{W(i)}{\sum_j W(j)}, j \in L,$$

is the stationary distribution of the Markov chain.

The invariant measure of Z^N

- Recall the one-step transition probabilities of Z^N:

$$P(K_i, K_j) \sim \exp\{-N\tilde{V}(K_i, K_j)\}.$$

- Let $W(K_i) = min_{g \in G(i)} \sum_{(m,n) \in g} \tilde{V}(K_m, K_n)$.
- By the Markov chain tree theorem, the the invariant measure of Z^N satisfies

$$\gamma_N(K_i) \sim \exp\{-N(W(i) - \min_j W(j))\}.$$

- Reconstruct $\wp_N$ from γ_N and show that

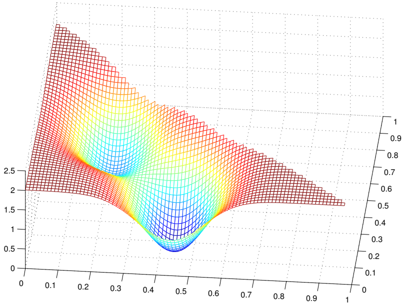
$$\wp_N(K_i) \sim \exp\{-N(W(i) - \min_j W(j))\}.$$

Large deviations of the invariant measure

Theorem

In the case of multiple equilibria, $\{\wp_N\}$ satisfies the LDP with rate function

$$V(\xi) = \min_{1\le i\le l}[W(K_i) + \tilde{V}(K_i,\xi)] - \min_{1\le i\le l} W(K_i)$$



Some applications of the LDP

- Exit times:
  - The mean exit time from K_i is of the order $\exp\{N\tilde{V}(K_i)\}$, where $\tilde{V}(K_i) = \min_j \tilde{V}(K_i, K_j)$.
- Mixing time of μ_N:
  - There is a constant $\Lambda > 0$ such that μ_N mixes well when the time is of the order $\exp\{N\Lambda\}$.
  - Proof via the exploration of equilibria. Mean passage times are of the order $\exp\{N\tilde{V}\}$, and has probability at least $\exp\{-N\varepsilon\}$.

Summary of Section 2

- A primer on large deviations.
- The process-level large deviations of the empirical measure process $\{\mu_N\}$.
  - Get the LDP for a non-interacting system using Sanov's theorem.
  - Use Varadhan's lemma to transfer it to $\{\mu_N\}$.
- Large deviations of the family of invariant measures $\{\wp_N\}$.
  - The unique attractor case: Identify the rate function from a recursion.
  - The multiple attractor case: Identify the values on the attractors.

Section 3

Variations - Two time-scales

Mean-Field Interacting Particle Systems: Limit Laws and Large Deviations

Section 3: Variations and Phenomena

SIGMETRICS/PERFORMANCE 2022

Outline of Section 3

- Variations:
  - A two time scale mean-field model.
  - Process-level large deviations of the empirical measure process.
- Phenomena:
  - A countable state mean-field model.
  - Large deviations of the family of invariant measures.
- Summary and some open questions.

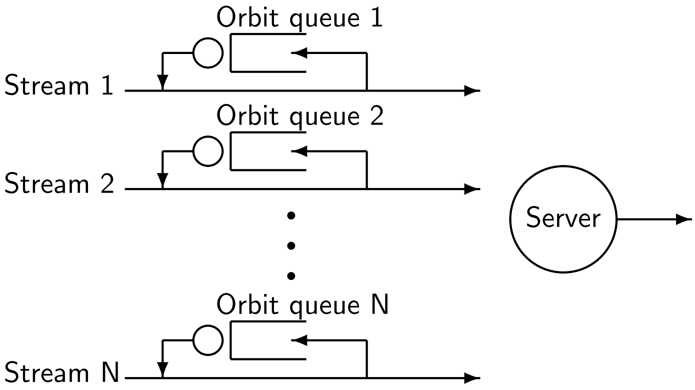
A two time scale mean-field model

- N particles and an environment.
- At time t,
  - The state of the nth particle is $X_n^N(t) \in \mathcal{Z}$;
  - The state of the environment is $Y_N(t) \in \mathcal{Y}$.
- Certain allowed transitions.
  - Particles: a directed graph $(\mathcal{Z}, \mathcal{E}_\mathcal{Z})$;
  - Environment: a directed graph $(\mathcal{Y}, \mathcal{E}_\mathcal{Y})$.
- Empirical measure of the system of particles at time t:

$$\mu_N(t) := \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(t)} \in \mathcal{M}_1(\mathcal{Z}).$$

- We are given functions $\lambda_{i,j}(\cdot, y)$, $(i,j) \in \mathcal{E}_\mathcal{Z}$, $y \in \mathcal{Y}$ and $\gamma_{y,y'}(\cdot)$, $(y, y') \in \mathcal{E}_\mathcal{Y}$ on $\mathcal{M}_1(\mathcal{Z})$.
- Markovian evolution at time t:
  - Particles: $i \to j$ at rate $\lambda_{i,j}(\mu_N(t), Y_N(t))$;
  - Environment: $y \to y'$ at rate $N\gamma_{y,y'}(\mu_N(t))$.

An example: Constant rate retrial systems



- N queues (particles), and a single server (environment).
- The server becomes busy at rate $N(\lambda + \alpha(1 - \mu_N(t)(0)))$.

A two time scale mean-field model

- (μ_N, Y_N) is a Markov process with the transition rates

$$(\xi, y) \to \begin{cases} (\xi, y') & \text{at rate } N\gamma_{y,y'}(\xi) \\ \left(\xi + \frac{\delta_j}{N} - \frac{\delta_i}{N}\right) & \text{at rate } N\xi(i)\lambda_{i,j}(\xi, y). \end{cases}$$

- A "fully coupled" two time scale process.
- Assumptions:
  - The graphs $(\mathcal{Z}, \mathcal{E}_{\mathcal{Z}})$ and $(\mathcal{Y}, \mathcal{E}_{\mathcal{Y}})$ are irreducible.
  - The functions $\lambda_{i,j}(\cdot, y)$ are Lipschitz continuous and $\inf_\xi \lambda_{i,j}(\xi, y) > 0$ for all $(i,j) \in \mathcal{E}_{\mathcal{Z}}$ and $y \in \mathcal{Y}$.
  - The functions $\gamma_{y,y'}(\cdot)$ are continuous and $\inf_\xi \gamma_{y,y'}(\xi) > 0$ for all $(y, y') \in \mathcal{E}_{\mathcal{Y}}$.

The occupation measure process

- Fix a time duration $T > 0$.
- View μ_N as a random element of $D([0,T],\mathcal{M}_1(\mathcal{Z}))$.
- Consider the occupation measure of the fast environment:

$$\theta_N(t)(\cdot) := \int_0^t 1_{\{Y_N(s)\in\cdot\}} ds,\ 0 \le t \le T.$$

- θ_N is a random element of $D_{\uparrow}([0,T],\mathcal{M}(\mathcal{Y}))$, the set of θ such that $\theta_t - \theta_s \in M(\mathcal{Y})$ and $\theta_t(\mathcal{Y}) = t$ for $0 \le s \le t \le T$.
- We can write θ as $\theta(dydt) = m_t(dy)dt$ where $m_t \in M_1(\mathcal{Y})$.
- We consider the process (μ_N, θ_N) with sample paths in $D([0,T],\mathcal{M}_1(\mathcal{Z})) \times D_{\uparrow}([0,T],\mathcal{M}(\mathcal{Y}))$.

The averaging principle

- Suppose we freeze $\mu_N(t)$ to be ξ. Then for large N,
  - The Y_N process would quickly equilibrate to π_ξ, the unique invariant probability measure of
    $$L_\xi g(y) := \sum_{y':(y,y')\in\mathcal{E}_\mathcal{Y}} (g(y') - g(y))\gamma_{y,y'}(\xi), y \in \mathcal{Y}.$$
  - For a particle, an (i,j) transition occurs at rate $\sum_{y\in\mathcal{Y}} \lambda_{i,j}(\xi,y)\pi_\xi(y) =: \bar{\lambda}_{i,j}(\xi,\pi_\xi)$.

Theorem (Bordenave et al. 2009)

Suppose that $\mu_N(0) \to \nu$ in $\mathcal{M}_1(\mathcal{Z})$. Then μ_N converges in probability, in $D([0,T],\mathcal{M}_1(\mathcal{Z}))$, to the solution to the ODE

$$\dot{\mu}_t = \bar{\Lambda}^T_{\mu_t,\pi_{\mu_t}}\mu_t,\ 0 \le t \le T,\ \mu_0 = \nu.$$

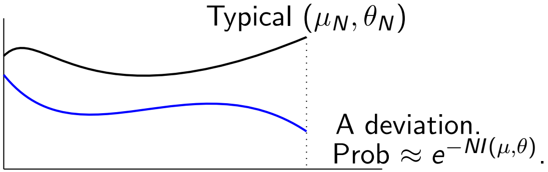
where $\bar{\Lambda}_{\mu_t,\pi_{\mu_t}}(i,j) = \bar{\lambda}_{i,j}(\mu_t,\pi_{\mu_t})$.

- μ_N is a small random perturbation of the above ODE. We study the large deviations of (μ_N,θ_N).

Main result

Theorem
Suppose that $\{\mu_N(0)\}_{N\geq 1}$ satisfies the LDP on $\mathcal{M}_1(\mathcal{Z})$ with rate function I_0. Then the sequence $\{(\mu_N(t), \theta_N(t)), 0 \leq t \leq T\}_{N\geq 1}$ satisfies the LDP on $D([0,T], \mathcal{M}_1(\mathcal{Z})) \times D_{\uparrow}([0,T], \mathcal{M}(\mathcal{Y}))$ with rate function

$$I(\mu, \theta) := I_0(\mu(0)) + J(\mu, \theta).$$



The rate function J

$$J(\mu,\theta) := \int_{[0,T]} \Bigg\{ \sup_{\alpha\in\mathbb{R}^{|\mathcal{Z}|}} \Bigg(\Big\langle \alpha, (\dot\mu_t - \bar\Lambda^T_{\mu_t,m_t}\mu_t) \Big\rangle - \sum_{(i,j)\in\mathcal{E}_{\mathcal{Z}}} \tau(\alpha(j)-\alpha(i))\bar\lambda_{i,j}(\mu_t,m_t)\mu_t(i) \Bigg) + \sup_{g\in\mathbb{R}^{|\mathcal{Y}|}} \sum_{y\in\mathcal{Y}} \Bigg(-L_{\mu_t}g(y) - \sum_{y':(y,y')\in\mathcal{E}_{\mathcal{Y}}} \tau(g(y')-g(y))\gamma_{y,y'}(\mu_t) \Bigg) m_t(y) \Bigg\} dt$$

whenever the mapping $[0,T] \ni t \mapsto \mu_t \in \mathcal{M}_1(\mathcal{Z})$ is absolutely continuous, where $\theta(dtdy) = m_t(dy)dt$, and $J(\mu,\theta) = +\infty$ otherwise.

- $\tau(u) = e^u - u - 1, u \in \mathbb{R}$.

Some remarks about the rate function

- $J(\mu, \theta) \geq 0$ with equality iff (μ, θ) satisfies the mean-field limit.
- Two parts. The mean-field part (slow component) and occupation measure part (fast component).
  - For the slow component, the average of the fast variable appears.
  - For the fast component, the slow variable is frozen.
- For occupation measure of Markov processes, the canonical form of the rate function is $\int_{[0,T]} \sup_{h>0} \sum_{\mathcal{Y}} -\frac{L_{\mu_t} h(y)}{h(y)} m_t(y) dt$ (Donsker and Varadhan, 1973). This can be obtained by taking $h = e^g$.

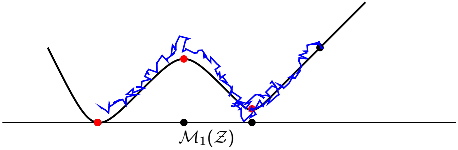
Large deviations of μ_N

Corollary

$\{\mu_N\}$ satisfies the LDP on $D([0,T],\mathcal{M}_1(\mathcal{Z}))$ with rate function

$$\mu \mapsto I_0(\mu_0) + \inf_{\theta} J(\mu,\theta).$$

- Follows from contraction principle since the mapping $(\mu,\theta)\mapsto\mu$ is continuous.
- Can quantify rare transitions.



Outline of the proof

- We use the method of stochastic exponentials (Pulahskii 2016, 1994).
- Show exponential tightness. This gives a subsequential LDP.
- Get a condition for any subsequential rate function (in terms of an exponential martingale).
- Identify the subsequential rate function on "nice" elements of the space.
- Extend to the whole space using suitable approximations.
- Unique identification any subsequential rate function (regardless of the subsequence) implies the LDP.

An exponential martingale

- If N_t is the unit rate Poisson point process, then $N_t - t$ is a martingale.
- Recall that

$$\tau(\alpha) = \log E(\exp\{\alpha(N_1 - 1)\}).$$

- One can verify that

$$\exp\{\alpha(N_t - t) - \tau(\alpha)t\}$$

  is a martingale for all α.
- We get a necessary condition for the subsequential rate function in terms of such exponential martingales.

Exponential tightness

Theorem

The sequence $\{(\mu_N(t), \theta_N(t)), t \in [0, T]\}_{N \geq 1}$ *is exponentially tight in* $D([0, T], M_1(\mathcal{Z})) \times D_{\uparrow}([0, T], M(\mathcal{Y}))$*, i.e., given any* $M > 0$*, there exists a compact set* K_M *such that*

$$\limsup_{N\to\infty} \frac{1}{N} \log P\left(\{(\mu_N(t), \theta_N(t)), 0 \leq t \leq T\} \notin K_M\right) \leq -M.$$

For $\beta > 0$ and $\alpha \in \mathbb{R}^{|\mathcal{Z}|}$, with $X_{N,t} = \langle \alpha, \mu_N(t) \rangle$,

$$\exp\Bigg\{ N\Bigg(\beta X_{N,t} - \beta X_{N,0} - \beta \int_0^t \Phi_{Y_{N,s}} f(\mu_{N,s}) ds - \int_0^t \sum_{(i,j)} \tau(\beta(\alpha(j) - \alpha(i)))\lambda_{i,j}(\mu_{N,s}, Y_{N,s})\mu_{N,s}(i) ds \Bigg) \Bigg\}, t \geq 0,$$

is an exponential martingale. Use Doob's inequality and a condition for exponential tightness in $D([0, T], \mathbb{R})$ (Puhalskii, 1994).

An equation for the subsequential rate function

- Let $\{(\mu_{N_k}, \theta_{N_k})\}_{k\geq 1}$ be a subsequence that satisfies the LDP with rate function $\tilde{I}$.
- Let $\alpha : [0, T] \times \mathcal{M}_1(\mathcal{Z}) \to \mathbb{R}^{|\mathcal{Z}|}$ and $g : [0, T] \times \mathcal{M}_1(\mathcal{Z}) \times \mathcal{Y} \to \mathbb{R}$ be bounded measurable, and continuous on $\mathcal{M}_1(\mathcal{Z})$.
- Define $U_t^{\alpha,g}(\mu, \theta)$ by

$$\int_{[0,t]} \Bigg\{ \langle \alpha_s(\mu_s), \dot{\mu}_s - \bar{\Lambda}^T_{\mu_s, m_s} \mu_s \rangle - \sum_{(i,j)} \tau(\alpha_s(\mu_s)(j) - \alpha_s(\mu_s)(i)) \bar{\lambda}_{i,j}(\mu_s, m_s) \mu_s(i) + \sum_y \Bigg(- L_{\mu_s} g_s(\mu_s, \cdot)(y) - \sum_{y:(y,y')\in\mathcal{E}_\mathcal{Y}} \tau(g_s(\mu_s, y') - g_s(\mu_s, y)) \gamma_{y,y'}(\mu_s) \Bigg) m_s(y) \Bigg\} ds.$$

An equation for the subsequential rate function

- We can show that, for each α and g,

$$\sup_{(\mu,\theta)\in\Gamma} \left(U_T^{\alpha,g}(\mu,\theta) - \tilde{I}(\mu,\theta)\right) = 0, \qquad (1)$$

  where Γ is the set of (μ,θ) such that $t \mapsto \mu_t$ absolutely continuous.
- On one hand, for a smaller class of α and g,

$$E \exp\{N U_T^{\alpha,g}(\mu_N, \theta_N) + V_T^g(\mu_N, Y_N)\} = 1,$$

  where V_T^g is $O(1)$ a.s.
- On the other hand, Varadhan's lemma implies that

$$\lim_{k\to\infty} \frac{1}{N_k} \log E \exp\{N_k U_T^{\alpha,g}(\mu_{N_k}, \theta_{N_k}) + V_T^g(\mu_{N_k}, Y_{N_k})\}$$
$$= \sup_{(\mu,\theta)} \left(U_T^{\alpha,g}(\mu,\theta) - \tilde{I}(\mu,\theta)\right)$$

  This can be extended to (1).
- Moreover, the supremum in (1) is attained.

A candidate rate function

- Recall that $\sup_{(\mu,\theta)\in\Gamma}(U_T^{\alpha,g}(\mu,\theta) - \tilde{I}(\mu,\theta)) = 0$.
- A natural candidate for the rate function

$$I^*(\mu,\theta) = \sup_{\alpha,g} U_T^{\alpha,g}(\mu,\theta).$$

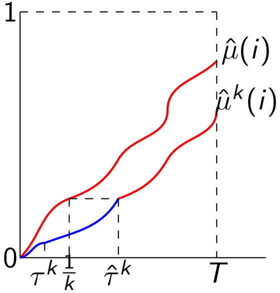
- It can be shown that $I^* = J$.
- Note that $\tilde{I} \geq I^*$ on Γ. Outside Γ, I^* can be shown to be $+\infty$.
- Goal: show that $\tilde{I} \leq I^*$ whenever $I^* < +\infty$. Once this is established, the LDP follows.

Identification of $\tilde{I}$ on "nice" elements

- Suppose $(\hat{\mu}, \hat{\theta})$ is such that $I^*(\hat{\mu}, \hat{\theta}) < +\infty$, and
  - $\inf_{t\in[0,T]} \min_{i\in\mathcal{Z}} \hat{\mu}_t(i) > 0$,
  - the mapping $[0, T] \ni t \mapsto \hat{\mu}_t \in \mathcal{M}_1(\mathcal{Z})$ is Lipschitz continuous,
  - $\inf_{t\in[0,T]} \min_{y\in\mathcal{Y}} \hat{m}_t(y) > 0$ where $\hat{\theta}(dydt) = \hat{m}_t(dy)dt$.
- Then, there exists $(\hat{\alpha}, \hat{g})$ that attains $\sup_{\alpha,g} U_T^{\alpha,g}(\hat{\mu}, \hat{\theta})$.
  - To show that $\hat{\alpha}$ and $\hat{g}$ are continuous on $\mathcal{M}_1(\mathcal{Z})$, we use the Berge's maximum theorem.
- With this $(\hat{\alpha}, \hat{g})$, get $(\tilde{\mu}, \tilde{\theta})$ that attains the supremum in $\sup_{(\mu,\theta)\in\Gamma}(U_T^{\hat{\alpha},\hat{g}}(\mu, \theta) - \tilde{I}(\mu, \theta)) = 0$.
- Hence, $I^*(\tilde{\mu}, \tilde{\theta}) \geq U_T^{\hat{\alpha},\hat{g}}(\tilde{\mu}, \tilde{\theta}) = \tilde{I}(\tilde{\mu}, \tilde{\theta})$.
- Since $I^* \leq \tilde{I}$, we get $I^*(\tilde{\mu}, \tilde{\theta}) = \tilde{I}(\tilde{\mu}, \tilde{\theta})$.
- Show that $(\tilde{\mu}, \tilde{\theta}) = (\hat{\mu}, \hat{\theta})$.
- It follows that $\tilde{I}(\hat{\mu}, \hat{\theta}) = I^*(\hat{\mu}, \hat{\theta})$.

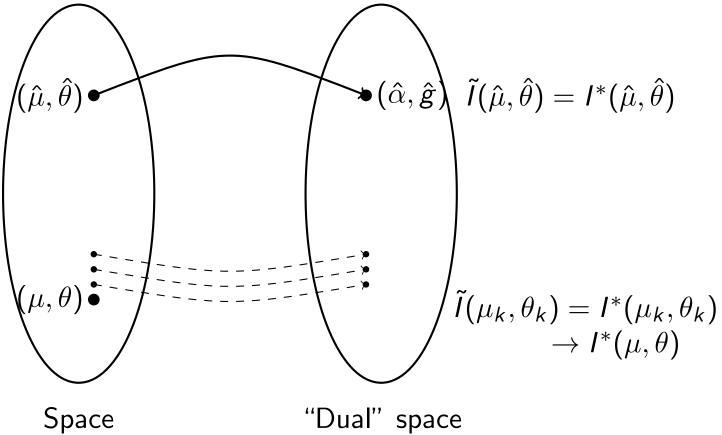
Approximation procedure

- For general elements $(\hat{\mu}, \hat{\theta})$, $(\hat{\alpha}, \hat{g})$ may not exist.
- Produce $(\hat{\mu}_k, \hat{\theta}_k)$ that are "nice", and satisfy
  - $(\hat{\mu}_k, \hat{\theta}_k) \to (\hat{\mu}, \hat{\theta})$ as $k \to \infty$,
  - $\tilde{I} = I^*$ on $(\hat{\mu}_k, \hat{\theta}_k)$ for all k,
  - $I^*(\hat{\mu}_k, \hat{\theta}_k) \to I^*(\hat{\mu}, \hat{\theta})$ as $k \to \infty$.
- It then follows that $\tilde{I} = I^*$ on $(\hat{\mu}, \hat{\theta})$.
- Relaxation of $\inf_{t \in [0,T]} \min_{i \in \mathcal{Z}} \hat{\mu}_t(i) > 0$:



- Other conditions are relaxed using suitable approximations. We finally get $\tilde{I} = I^*$ for all elements.

Summary of the proof

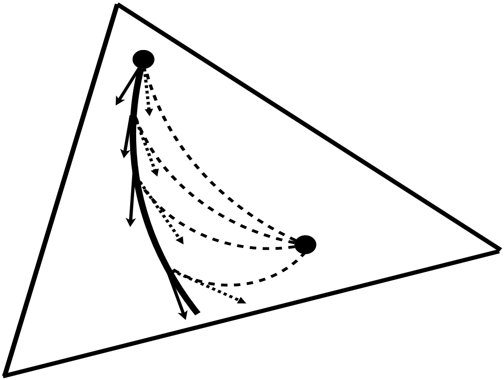


- For "nice" elements of $D([0, T], \mathcal{M}_1(\mathcal{Z})) \times D_\uparrow([0, T], \mathcal{M}(\mathcal{Y}))$, we show that $\tilde{I} = I^*$ (convex analysis, variational problems).
- Approximate general elements using "nice" elements and pass to the limit (parametric continuity of optimisation problems, dominated convergence).

Section 4

Variations - Phenomena in the infinite state space case

The running cost of following a trajectory $\phi(\cdot)$



- At each time t, if the current state is $\phi(t)$, the natural tendency is to go along the tangent $\Lambda(\phi(t))^T\phi(t)$.
- To follow $\phi(t)$ however, the system needs to work against the McKean-Vlasov gradient and move along the tangent $\dot{\phi}(t)$.
- $L(\phi(t), \dot{\phi}(t))$.

Guessing the running cost

- Write $\dot{\phi}(t) = G(t)^T \phi(t)$.
- By decoupling, each node's state is iid $\phi(t)$.
- Natural tendency for the $N\phi(t)(i)$ nodes in state i is to have $i \rightsquigarrow j$ at current (instantaneous) rate $\lambda_{i,j}(\phi(t))$.
- But to move along $\phi(t)$ they must have an instantaneous rate of $G_{i,j}(t)$.
- The $N\phi(t)(i)$ Bernoulli($p = \lambda_{i,j}(t)\ dt$) random variables must have a large deviation and must have an empirical measure close to $(q = G_{i,j}(t)\ dt)$. By Sanov's theorem, the negative exponent is:

$$N\phi(t)(i)D(q||p) \cong N\phi(t)(i)(q \log \frac{q}{p} - q + p)$$

- Sum over i and j and integrate over $[0, T]$ to get the action functional:

$$\int_0^T L(\phi(t), \dot{\phi}(t))\ dt.$$

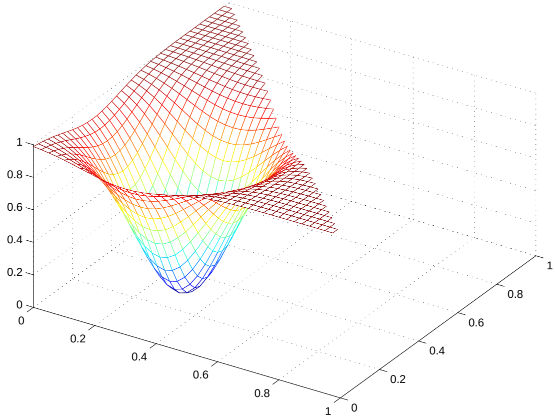
The case of a globally asymptotically stable equilibrium ξ^*

Theorem
$V(\xi)$ is given by

$$V(\xi) = \inf \left\{ \int_0^T L(\phi(t), \dot{\phi}(t)) \, dt \mid \phi(0) = \xi^*, \phi(T) = \xi, T \in (0, \infty) \right\}.$$

- Any deviation that puts the system at ξ must have started its effort from ξ^*.
- $V(\xi^*) = 0$.

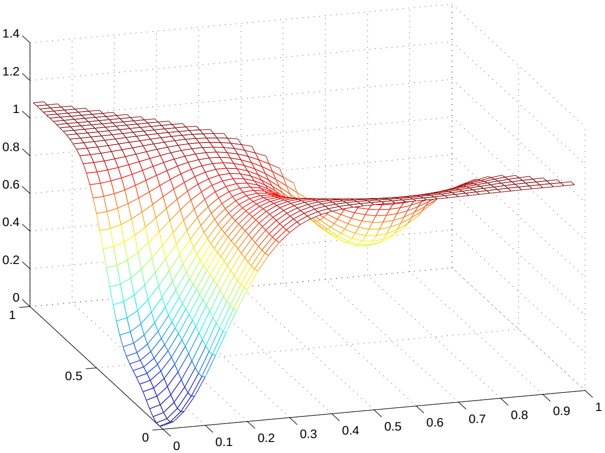
The path to ξ



Can specify not only exponent $V(\xi)$ of the probability, but also the path.

Any deviation that puts the system near q must have started from ξ^*, and must have taken the least cost path.

When there are multiple stable limit sets



The case of two stable equilibria is easy to describe.

- V_{12} = cost of moving from ξ_1^* to ξ_2^*.
- V_{21} = cost of the reverse move.
- If $V_{12} > V_{21}$, then $v_1 = 0$ and $v_2 = V_{12} - V_{21}$.

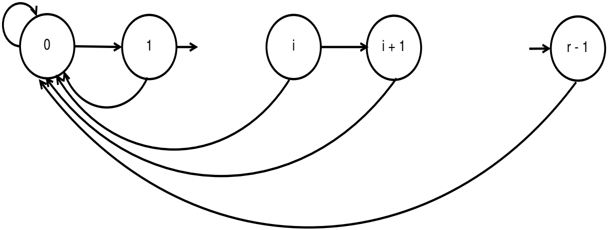
When there are multiple stable limit sets

Theorem
$V(\xi)$ is given by

$$V(\xi) = \inf_i \left\{ v_i + \int_0^T L(\phi(t), \dot{\phi}(t))\, dt \mid \phi(0) = \xi_i^*, \phi(T) = \xi, T \in (0, \infty) \right\}.$$

- Start from the global minimum ξ_1^* and move to the attractor in the basin in which ξ lies along the least cost path.
- Then move to ξ along the least cost path.

Infinite state space



- Now $r = \infty$
- Forward rate λ_f, backward rate λ_b. Let ξ^* be the invariant measure.
- $X_n^{(N)}(\infty) \sim \xi^*$
- $\xi^*(i) = (1-\rho)\rho^i, \quad i \geq 0$, where $\rho = \frac{\lambda_f}{\lambda_f + \lambda_b}$.

The "interacting particle system", LDP, and the rate function

- For explicit calculations, assume that the queues are noninteracting (i.e., each evolves independently).
- We are interested in invariant measure for the empirical measure.
- The invariant measure is just the law of $\mu_N(\infty) = \frac{1}{N}\sum_{n=1}^{N} \delta_{X_n^{(N)}(\infty)}$
- (Sanov) The $\mu_N(\infty)$ sequence satisfies the LDP with rate function given by relative entropy $I(\cdot\|\xi^*)$.

What are "reachable" points at stationarity?

- Let $\iota(i) = i$.

- $I(\xi \| \xi^*)$ is finite if and only if $\langle \xi, \iota \rangle < \infty$.

- Define $\vartheta(i) = i \log i$. There are points ξ for which $\langle \xi, \iota \rangle < \infty$, but $\langle \xi, \vartheta \rangle = \infty$. Mass is sufficiently spread out, since $I(\xi, \xi^*)$ is finite, they are still reachable at stationarity.

Quasipotential

- Define the quasipotential as before.

$$
\begin{aligned}
V(\xi) &= \inf\left\{\int_0^T L(\phi(t),\dot{\phi}(t))\,dt \mid \phi(0)=\xi^*, \phi(T)=\xi, T\in(0,\infty)\right\} \\
&\geq \inf_T \sup_{f\in C_0^1([0,T]\times\mathcal{Z}} \left\{\langle\phi_T, f_T\rangle - \langle\phi_0, f_0\rangle - \int_0^T \langle\phi_u, \partial_u f_u\rangle du \right. \\
&\quad \left. - \int_0^T \langle\phi_u, \Lambda_{\phi_u} f_u\rangle du - \int_0^T \sum_{i,j} \tau(f_u(j)-f_u(i))\lambda_{i,j}(\phi_u)\phi_u(i)du \right\}
\end{aligned}
$$

- Last two terms simplify to $\int_0^T \sum_{i,j} \exp\{f_u(j)-f_u(i)\}\lambda_{i,j}(\phi_u)\phi_u(i)du$
- Strategy
  - Choose $f_n = \vartheta(Hat(0,n,2n))$. This is like $\vartheta(n)$ up to n.
  - Then $f_n(j) - f_n(i) \leq 1 + \log(i+1)$ for the edges in the graph.
  - Last two terms $\propto \langle\phi_u, \iota\rangle$ which integrates to a finite value.
  - Then let $f_n \to \vartheta$ as $n\to\infty$.
  - Then $\langle\xi,\vartheta\rangle = \infty \Rightarrow V(\xi) = \infty$.

Infinite state space

Theorem

The rate function for the invariant measure is the relative entropy $I(\cdot\|\xi^*)$, and this is not equal to the quasipotential V.

- Take a ξ whose mean is finite but the slightly larger $i \log i$ moment is infinite.

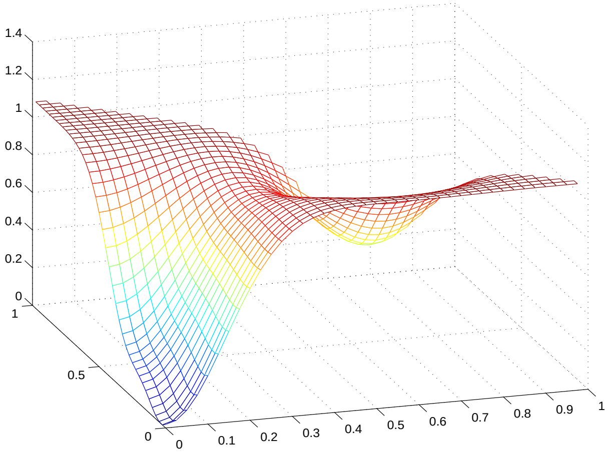
- V comes from a finite horizon perspective. There are barriers that are too difficult to cross in any finite time horizon, but in the stationary regime these can be crossed leading to a finite rate function at these points.

- A partial answer

Theorem

If $\lambda_{i,i+1}(\cdot) = \Theta(1/(i+1))$, then the rate function for the invariant measure is indeed governed by the quasipotential.

The take-away picture



$V_{1\to 2} > V_{2\to 1}$