Metastability phenomenon: Large deviations in the stationary regime

Sarath A Y

ECE Department, Indian Institute of Science

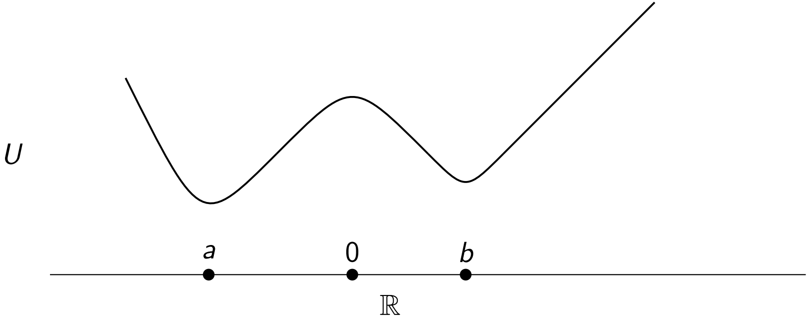
28 July 2020

Part 1: Metastability

Example 1 – Double well potential

- Consider the ODE

$$dX_t = -U'(X_t)dt.$$

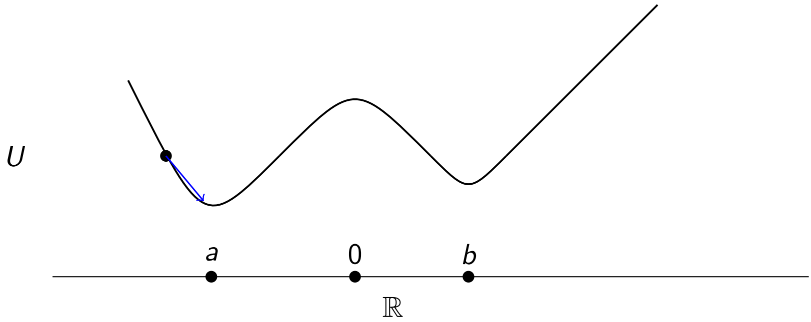


- a and b are stable equilibria. 0 is an unstable equilibrium.

Example 1 – Double well potential

- Consider the ODE

$$dX_t = -U'(X_t)dt.$$

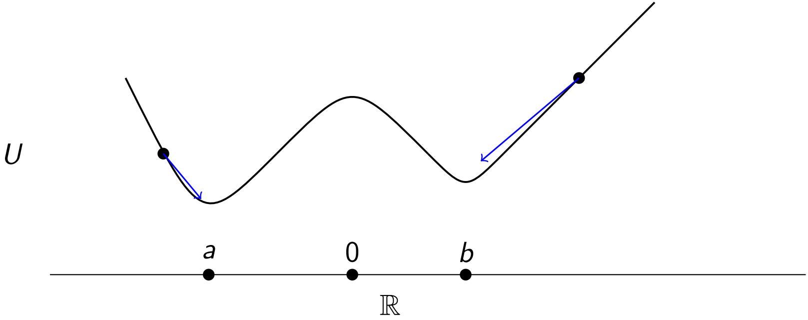


- a and b are stable equilibria. 0 is an unstable equilibrium.

Example 1 – Double well potential

- Consider the ODE

$$dX_t = -U'(X_t)dt.$$

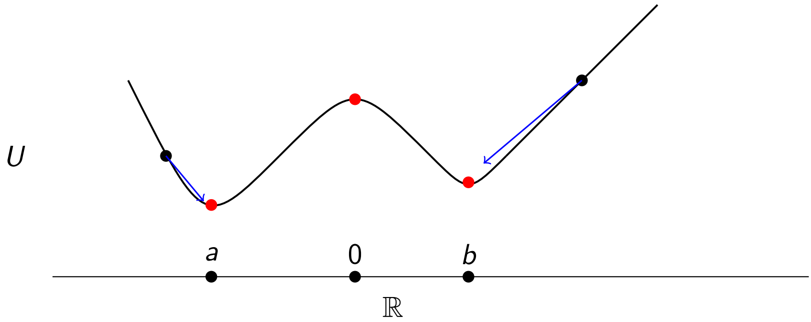


- a and b are stable equilibria. 0 is an unstable equilibrium.

Example 1 – Double well potential

- Consider the ODE

$$dX_t = -U'(X_t)dt.$$



- a and b are stable equilibria. 0 is an unstable equilibrium.

Example 1 with Gaussian noise

- Consider the perturbed ODE

$$dX_t^\varepsilon = -U'(X_t^\varepsilon)dt + \varepsilon dB_t$$

where dB_t is the white noise ($B_t \sim N(0, t)$, i.s.i.).

Example 1 with Gaussian noise

- Consider the perturbed ODE

$$dX_t^\varepsilon = -U'(X_t^\varepsilon)dt + \varepsilon dB_t$$

  where dB_t is the white noise ($B_t \sim N(0,t)$, i.s.i.).
- Turns out that X^ε is a Markov process.

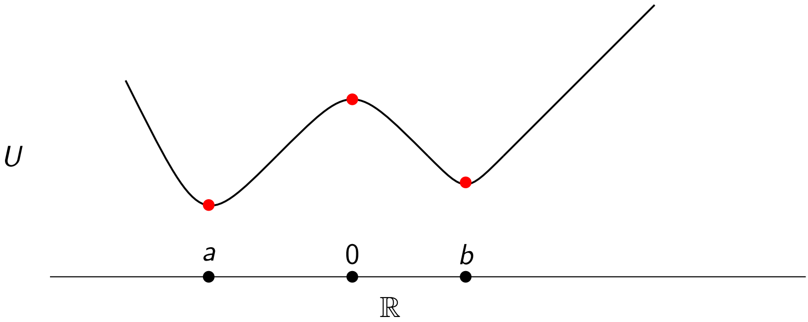
Example 1 with Gaussian noise

- Consider the perturbed ODE

$$dX_t^\varepsilon = -U'(X_t^\varepsilon)dt + \varepsilon dB_t$$

where dB_t is the white noise ($B_t \sim N(0, t)$, i.s.i.).

- Turns out that X^ε is a Markov process.



- Metastability phenomenon: Very different behaviour over different time scales.

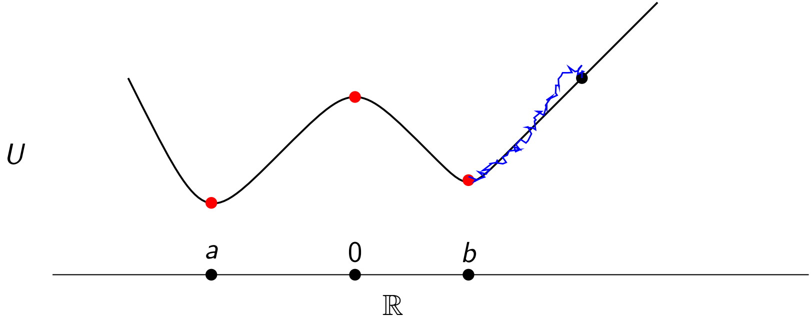
Example 1 with Gaussian noise

- Consider the perturbed ODE

$$dX_t^\varepsilon = -U'(X_t^\varepsilon)dt + \varepsilon dB_t$$

where dB_t is the white noise ($B_t \sim N(0,t)$, i.s.i.).

- Turns out that X^ε is a Markov process.



- Metastability phenomenon: Very different behaviour over different time scales.

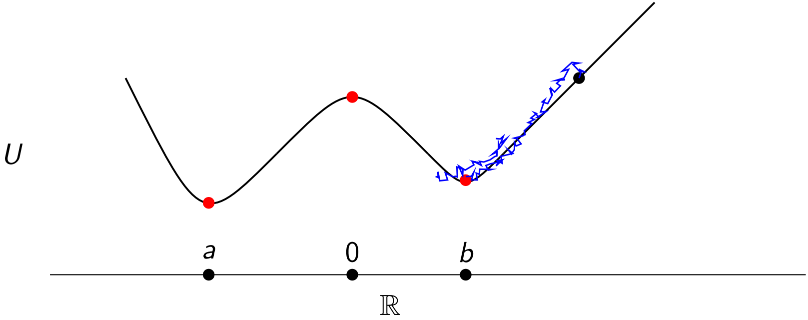
Example 1 with Gaussian noise

- Consider the perturbed ODE

$$dX_t^\varepsilon = -U'(X_t^\varepsilon)dt + \varepsilon dB_t$$

where dB_t is the white noise ($B_t \sim N(0,t)$, i.s.i.).

- Turns out that X^ε is a Markov process.



- Metastability phenomenon: Very different behaviour over different time scales.

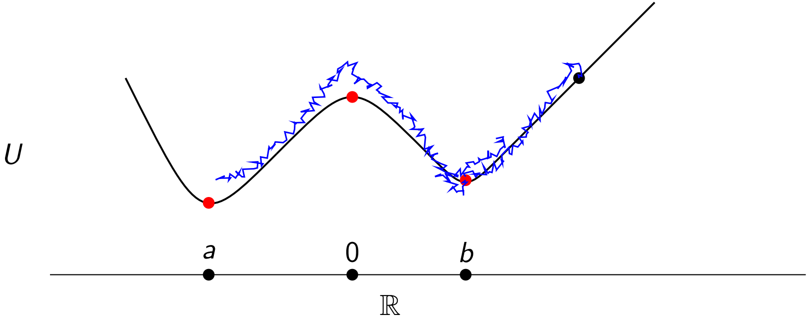
Example 1 with Gaussian noise

- Consider the perturbed ODE

$$dX_t^\varepsilon = -U'(X_t^\varepsilon)dt + \varepsilon dB_t$$

where dB_t is the white noise ($B_t \sim N(0, t)$, i.s.i.).

- Turns out that X^ε is a Markov process.



- Metastability phenomenon: Very different behaviour over different time scales.

Example 2 – A mean-field model

- N particles. State space: a finite set $\mathcal{Z}$.
- State of the nth particle at time t is $X_n^N(t) \in \mathcal{Z}$.

Example 2 – A mean-field model

- N particles. State space: a finite set $\mathcal{Z}$.
- State of the nth particle at time t is $X_n^N(t) \in \mathcal{Z}$.
- Certain allowed transitions: specified by a directed graph $(\mathcal{Z}, \mathcal{E})$

Example 2 – A mean-field model

- N particles. State space: a finite set $\mathcal{Z}$.
- State of the nth particle at time t is $X_n^N(t) \in \mathcal{Z}$.
- Certain allowed transitions: specified by a directed graph $(\mathcal{Z}, \mathcal{E})$
- Empirical measure at time t

$$\mu_N(t) = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(t)} \in M_1(\mathcal{Z}).$$

Example 2 – A mean-field model

- N particles. State space: a finite set $\mathcal{Z}$.
- State of the nth particle at time t is $X_n^N(t) \in \mathcal{Z}$.
- Certain allowed transitions: specified by a directed graph $(\mathcal{Z}, \mathcal{E})$
- Empirical measure at time t

$$\mu_N(t) = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(t)} \in M_1(\mathcal{Z}).$$

- For each $(z, z') \in \mathcal{E}$, we have a function $\lambda_{z,z'} : M_1(\mathcal{Z}) \to [0, +\infty)$.
- Particle transitions: at time t, a $z \to z'$ transition occurs at rate $\lambda_{z,z'}(\mu_N(t))$. Mean-field interaction.

Example 2 – A mean-field model

- N particles. State space: a finite set $\mathcal{Z}$.
- State of the nth particle at time t is $X_n^N(t) \in \mathcal{Z}$.
- Certain allowed transitions: specified by a directed graph $(\mathcal{Z}, \mathcal{E})$
- Empirical measure at time t

$$\mu_N(t) = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n^N(t)} \in M_1(\mathcal{Z}).$$

- For each $(z, z') \in \mathcal{E}$, we have a function $\lambda_{z,z'} : M_1(\mathcal{Z}) \to [0, +\infty)$.
- Particle transitions: at time t, a $z \to z'$ transition occurs at rate $\lambda_{z,z'}(\mu_N(t))$. Mean-field interaction.
- $\{(X_n^N(t), 1 \le n \le N), t \ge 0\}$ is a Markov process on $\mathcal{Z}^N$.

The empirical measure process μ_N

- $\{\mu_N(t), t \geq 0\}$ is a Markov process on $M_1(\mathcal{Z})$ with infinitesimal generator

$$L^N f(\xi) = \sum_{(z,z')\in\mathcal{E}} N\xi(z)\lambda_{z,z'}(\xi)\left[f\left(\xi + \frac{\delta_{z'}}{N} - \frac{\delta_z}{N}\right) - f(\xi)\right].$$

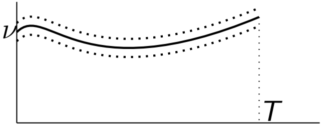
The empirical measure process μ_N

- $\{\mu_N(t), t \geq 0\}$ is a Markov process on $M_1(\mathcal{Z})$ with infinitesimal generator

$$L^N f(\xi) = \sum_{(z,z')\in\mathcal{E}} N\xi(z)\lambda_{z,z'}(\xi)\left[f\left(\xi + \frac{\delta_{z'}}{N} - \frac{\delta_z}{N}\right) - f(\xi)\right].$$

- Typical behaviour of μ_N (mean-field limit):
  Let $\mu_N(0) \to \nu$ weakly as $N \to \infty$. Then $\{\mu_N(t), 0 \leq t \leq T\}$, w.h.p., is "close to" to the solution to the McKean-Vlasov equation:

$$\dot{\mu}_t = \Lambda^*_{\mu_t}\mu_t, \; \mu_0 = \nu.$$



- Thus, μ_N is an ODE+"noise". Metastability occurs here as well.

Interaction in WiFi networks

- N nodes accessing a common wireless medium.
- Interaction among nodes via the distributed MAC protocol.

Interaction in WiFi networks

- N nodes accessing a common wireless medium.
- Interaction among nodes via the distributed MAC protocol.
- State $X_n^N(t)$ represents aggressiveness of packet transmission.
-

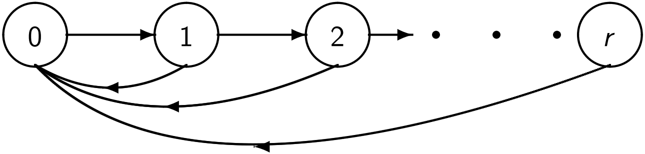


Figure: Set of allowed transitions in WiFi example

Interaction in WiFi networks

- N nodes accessing a common wireless medium.
- Interaction among nodes via the distributed MAC protocol.
- State $X_n^N(t)$ represents aggressiveness of packet transmission.
-

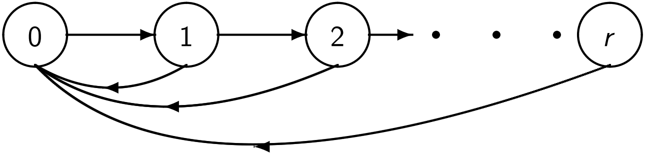


Figure: Set of allowed transitions in WiFi example

- State evolution:
  - Becomes less aggressive after a collision.
  - Moves to the most aggressive state after a successful packet transmission.

A sample path of μ_N in WiFi example

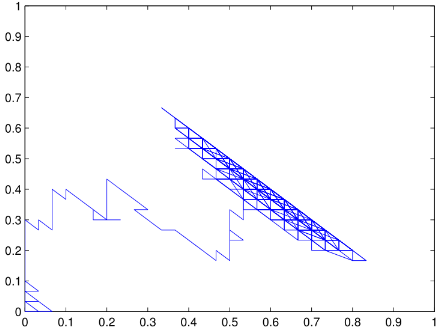


Figure: Evolution of states in a WiFi network under the MAC protocol

- Metastability phenomenon: Multiple stable regions in the system. Transition between two stable regions occur over large time durations.

Random perturbations of dynamical systems

- Example 1 - Double well potential
  - Perturbation of an ODE, “Gaussian” noise.
  - Markov process, a.s.-continuous sample paths
  - State space: $\mathbb{R}$
- Example 2 - Mean-field model
  - Perturbation of the McKean-Vlasov equation, “Poissonian noise”.
  - Markov process, with jumps
  - State space: $M_1(\mathcal{Z})$
- Abstractly,
  - A dynamical system, taking values in some space. $\mathbb{R}, M_1(\mathcal{Z}), C([0,1])$, etc.
  - A small random perturbation of this dynamical system. Mostly, Gaussian or Poissonian noise. This generates a Markov process.
  - As noise becomes small, the process converges to its typical behaviour.
  - Goal of this talk: Explain the metastability phenomenon of randomly perturbed dynamical system.

Back to Example 1 – A more concrete goal

- The potential U explains metastability phenomenon.
- It turns out that X^ε is reversible and possesses a unique invariant probability measure

$$\wp^\varepsilon(dx) = \text{const} \times \exp\left\{-\frac{1}{\varepsilon^2} U(x) dx\right\}.$$

Back to Example 1 – A more concrete goal

- The potential U explains metastability phenomenon.
- It turns out that X^ε is reversible and possesses a unique invariant probability measure

$$\wp^\varepsilon(dx) = \text{const} \times \exp\left\{-\frac{1}{\varepsilon^2} U(x)dx\right\}.$$

- In particular, for "nice sets" $A \subset \mathbb{R}$, one anticipates

$$\lim_{\varepsilon\to 0} \varepsilon^2 \log \wp^\varepsilon(A) = -\inf_{x\in A} U(x).$$

Back to Example 1 – A more concrete goal

- The potential U explains metastability phenomenon.
- It turns out that X^ε is reversible and possesses a unique invariant probability measure

$$\wp^\varepsilon(dx) = \text{const} \times \exp\left\{-\frac{1}{\varepsilon^2} U(x) dx\right\}.$$

- In particular, for "nice sets" $A \subset \mathbb{R}$, one anticipates

$$\lim_{\varepsilon \to 0} \varepsilon^2 \log \wp^\varepsilon(A) = -\inf_{x \in A} U(x).$$

- But, there is no closed for expression for $\wp^\varepsilon$ in most cases!

Back to Example 1 – A more concrete goal

- The potential U explains metastability phenomenon.
- It turns out that X^ε is reversible and possesses a unique invariant probability measure

$$\wp^\varepsilon(dx) = \text{const} \times \exp\left\{-\frac{1}{\varepsilon^2}U(x)dx\right\}.$$

- In particular, for "nice sets" $A \subset \mathbb{R}$, one anticipates

$$\lim_{\varepsilon\to 0} \varepsilon^2 \log \wp^\varepsilon(A) = -\inf_{x\in A} U(x).$$

- But, there is no closed for expression for $\wp^\varepsilon$ in most cases!
  - The WiFi example has no closed form expression, even when the nodes do not interact.

Back to Example 1 – A more concrete goal

- The potential U explains metastability phenomenon.
- It turns out that X^ε is reversible and possesses a unique invariant probability measure

$$\wp^\varepsilon(dx) = \text{const} \times \exp\left\{-\frac{1}{\varepsilon^2} U(x)dx\right\}.$$

- In particular, for "nice sets" $A \subset \mathbb{R}$, one anticipates

$$\lim_{\varepsilon \to 0} \varepsilon^2 \log \wp^\varepsilon(A) = -\inf_{x \in A} U(x).$$

- But, there is no closed for expression for $\wp^\varepsilon$ in most cases!
  - The WiFi example has no closed form expression, even when the nodes do not interact.
- We want to identify a suitable "U" and obtain asymptotics of the above form.
  - Other questions: Asymptotics of exit times, mixing time etc. (not in this talk).

Part 2: Large deviations

Large deviations

- S: a metric space. $\{X_N\}_{N\geq 1}$ is a sequence of S-valued random variables.
- Roughly, $P(X_N \in A) \sim \exp\{-NI_A\}$ where $I_A = \inf_{x\in A} I(x)$.

Large deviations

- S: a metric space. $\{X_N\}_{N\geq 1}$ is a sequence of S-valued random variables.
- Roughly, $P(X_N \in A) \sim \exp\{-NI_A\}$ where $I_A = \inf_{x\in A} I(x)$.

Definition (Varadhan, 1966)

$\{X_N\}_{N\geq 1}$ is said to satisfy the large deviation principle (LDP) with rate function $I : S \to [0, +\infty]$ if

Large deviations

- S: a metric space. $\{X_N\}_{N\geq 1}$ is a sequence of S-valued random variables.
- Roughly, $P(X_N \in A) \sim \exp\{-NI_A\}$ where $I_A = \inf_{x\in A} I(x)$.

Definition (Varadhan, 1966)

$\{X_N\}_{N\geq 1}$ is said to satisfy the large deviation principle (LDP) with rate function $I : S \to [0, +\infty]$ if

- for each $M > 0$, $\{x \in S : I(x) \leq M\}$ is a compact subset of S,

Large deviations

- S: a metric space. $\{X_N\}_{N\geq 1}$ is a sequence of S-valued random variables.
- Roughly, $P(X_N \in A) \sim \exp\{-NI_A\}$ where $I_A = \inf_{x\in A} I(x)$.

Definition (Varadhan, 1966)

$\{X_N\}_{N\geq 1}$ is said to satisfy the large deviation principle (LDP) with rate function $I : S \to [0, +\infty]$ if

- for each $M > 0$, $\{x \in S : I(x) \leq M\}$ is a compact subset of S,
- for each open set $G \subset S$

$$\liminf_{N\to\infty} \frac{1}{N} \log P(X_N \in G) \geq - \inf_{x\in G} I(x);$$

Large deviations

- S: a metric space. $\{X_N\}_{N\geq 1}$ is a sequence of S-valued random variables.
- Roughly, $P(X_N \in A) \sim \exp\{-NI_A\}$ where $I_A = \inf_{x\in A} I(x)$.

Definition (Varadhan, 1966)

$\{X_N\}_{N\geq 1}$ is said to satisfy the large deviation principle (LDP) with rate function $I : S \to [0, +\infty]$ if

- for each $M > 0$, $\{x \in S : I(x) \leq M\}$ is a compact subset of S,
- for each open set $G \subset S$

$$\liminf_{N\to\infty} \frac{1}{N} \log P(X_N \in G) \geq -\inf_{x\in G} I(x);$$

- for each closed set $F \subset S$

$$\limsup_{N\to\infty} \frac{1}{N} \log P(X_N \in F) \leq -\inf_{x\in F} I(x).$$

Large deviations: An equivalent formulation

Definition (Freidlin and Wentzell, $\sim$70s)

$\{X_N\}$ is said to satisfy the LDP with rate function $I : S \to [0, \infty]$ if

Large deviations: An equivalent formulation

Definition (Freidlin and Wentzell, $\sim$70s)

$\{X_N\}$ is said to satisfy the LDP with rate function $I : S \to [0, \infty]$ if

- for any $s \geq 0$, $\Phi(s) = \{x \in S : I(x) \leq s\}$ is a compact subset of S;

Large deviations: An equivalent formulation

Definition (Freidlin and Wentzell, ∼70s)

$\{X_N\}$ is said to satisfy the LDP with rate function $I : S \to [0, \infty]$ if

- for any $s \geq 0$, $\Phi(s) = \{x \in S : I(x) \leq s\}$ is a compact subset of S;
- for any $\gamma > 0, \delta > 0, s > 0$ there exists $N_0 \geq 1$ such that

$$P(d(X_N, x) < \delta) \geq \exp\{-N(I(x) + \gamma)\}$$

for any $x \in \Phi(s)$ and $N \geq N_0$;

Large deviations: An equivalent formulation

Definition (Freidlin and Wentzell, ∼70s)

$\{X_N\}$ is said to satisfy the LDP with rate function $I : S \to [0, \infty]$ if

- for any $s \geq 0$, $\Phi(s) = \{x \in S : I(x) \leq s\}$ is a compact subset of S;
- for any $\gamma > 0, \delta > 0, s > 0$ there exists $N_0 \geq 1$ such that

  $$P(d(X_N, x) < \delta) \geq \exp\{-N(I(x) + \gamma)\}$$

  for any $x \in \Phi(s)$ and $N \geq N_0$;
- for any $\gamma > 0, \delta > 0, s_0 > 0$ there exists $N_0 \geq 1$ such that

  $$P(d(X_N, \Phi(s)) \geq \delta) \leq \exp\{-N(s - \gamma)\}$$

  for any $s \leq s_0$ and $N \geq N_0$.

More suited to study the stationary regime. We'll come back later.

Large deviations: An example

- Let S be a complete and separable metric space. ν is a probability measure on S.
  - Example: $S = \mathbb{R}$, ν is the standard Gaussian.

Large deviations: An example

- Let S be a complete and separable metric space. ν is a probability measure on S.
  - Example: $S = \mathbb{R}$, ν is the standard Gaussian.
- Let $X_1, X_2, \ldots, X_N$ be i.i.d. ν.

Large deviations: An example

- Let S be a complete and separable metric space. ν is a probability measure on S.
  - Example: $S = \mathbb{R}$, ν is the standard Gaussian.
- Let $X_1, X_2, \ldots, X_N$ be i.i.d. ν.
- Anticipate that

$$M_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n} \to \nu$$

weakly as $N \to \infty$, in probability.

Large deviations: An example

- Let S be a complete and separable metric space. ν is a probability measure on S.
  - Example: $S = \mathbb{R}$, ν is the standard Gaussian.
- Let $X_1, X_2, \ldots, X_N$ be i.i.d. ν.
- Anticipate that

$$M_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n} \to \nu$$

  weakly as $N \to \infty$, in probability.
- M_N is an $M_1(S)$-valued random variable.

Large deviations: An example

- Let S be a complete and separable metric space. ν is a probability measure on S.
  - Example: $S = \mathbb{R}$, ν is the standard Gaussian.
- Let $X_1, X_2, \dots, X_N$ be i.i.d. ν.
- Anticipate that

$$M_N = \frac{1}{N} \sum_{n=1}^{N} \delta_{X_n} \to \nu$$

  weakly as $N \to \infty$, in probability.
- M_N is an $M_1(S)$-valued random variable.

Theorem (Sanov)

$\{M_N\}$ satisfies the LDP on $M_1(S)$ with rate function $D(\cdot \| \nu)$.

- In particular, $P\left(M_N \in A\right) \sim \exp\{-N \inf_{\mu \in A} D(\mu \| \nu)\}$.

Large deviations: Contraction principle

- A means by which we can obtain new LDPs from old ones.

Large deviations: Contraction principle

- A means by which we can obtain new LDPs from old ones.
- S, T are metric spaces. $f : S \to T$ is continuous.
- $\{X_N\}$s are S- valued random variables. Define $Y_N = f(X_N)$.

Large deviations: Contraction principle

- A means by which we can obtain new LDPs from old ones.
- S, T are metric spaces. $f : S \to T$ is continuous.
- $\{X_N\}$s are S- valued random variables. Define $Y_N = f(X_N)$.
- If $\{X_N\}$ satisfies the LDP with rate function I, then $\{Y_N\}$ satisfies the LDP with rate function

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

Large deviations: Contraction principle

- A means by which we can obtain new LDPs from old ones.
- S, T are metric spaces. $f : S \to T$ is continuous.
- $\{X_N\}$s are S- valued random variables. Define $Y_N = f(X_N)$.
- If $\{X_N\}$ satisfies the LDP with rate function I, then $\{Y_N\}$ satisfies the LDP with rate function

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

  - Compactness of level sets: $\{y \in T : J(y) \leq M\} = f(\{x \in S : I(x) \leq M\})$.
  - Upper and lower bounds: $P(Y_N \in A) = P(X_N \in f^{-1}(A))$

Back to our goal

- Recall that we have to find a suitable "U" that governs asymptotics of the form

$$\lim_{\varepsilon \to 0} \varepsilon^2 \log \wp^{\varepsilon}(A) = - \inf_{x \in A} U(x).$$

Back to our goal

- Recall that we have to find a suitable "U" that governs asymptotics of the form

$$\lim_{\varepsilon \to 0} \varepsilon^2 \log \wp^\varepsilon(A) = -\inf_{x \in A} U(x).$$

- We only know that $\wp^\varepsilon$ is invariant for X^ε.

Back to our goal

- Recall that we have to find a suitable "U" that governs asymptotics of the form

$$\lim_{\varepsilon \to 0} \varepsilon^2 \log \wp^\varepsilon(A) = - \inf_{x \in A} U(x).$$

- We only know that $\wp^\varepsilon$ is invariant for X^ε.
- Strategy is to study large deviations of the trajectories and then move to stationarity.

Back to Example 1 – Large deviations in path-space

- Consider the trajectory $(X_t^\varepsilon, t \in [0, T])$ as a $C([0, T], \mathbb{R})$-valued random variable.

Back to Example 1 – Large deviations in path-space

- Consider the trajectory $(X_t^\varepsilon, t \in [0, T])$ as a $C([0, T], \mathbb{R})$-valued random variable.
- Anticipate that, when $X_0^\varepsilon \to x$,

$$(X_t^\varepsilon, t \in [0, T]) \to (X_t^0, t \in [0, T])$$

in $C([0, T], \mathbb{R})$, in probability, as $\varepsilon \to 0$ where X^0 solves the ODE

$$dX_t^0 = -U'(X_t^0)dt,\ X_0^0 = x.$$

Back to Example 1 – Large deviations in path-space

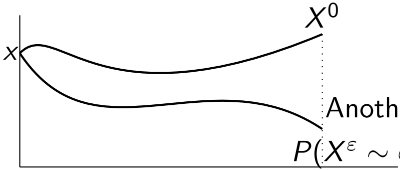
- Consider the trajectory $(X_t^\varepsilon, t \in [0,T])$ as a $C([0,T],\mathbb{R})$-valued random variable.
- Anticipate that, when $X_0^\varepsilon \to x$,

$$(X_t^\varepsilon, t \in [0,T]) \to (X_t^0, t \in [0,T])$$

  in $C([0,T],\mathbb{R})$, in probability, as $\varepsilon \to 0$ where X^0 solves the ODE

$$dX_t^0 = -U'(X_t^0)dt,\ X_0^0 = x.$$

- LDP for X^ε. A functional $S_{[0,T]}$ on $C([0,T],\mathbb{R})$ so that



Another path φ.

$P(X^\varepsilon \sim \varphi) \sim e^{-\frac{1}{\varepsilon^2} S_{[0,T]}(\varphi)}$.

Back to Example 1 – Large deviations in path-space

Theorem

Let $X_0^\varepsilon \to x$.

Back to Example 1 – Large deviations in path-space

Theorem

Let $X_0^\varepsilon \to x$*. Then* $(X_t^\varepsilon, t \in [0, T])$ *satisfies the LDP on* $C([0, T], \mathbb{R})$ *with rate function* $S_{[0,T]}$ *defined as follows.*

Back to Example 1 – Large deviations in path-space

Theorem

Let $X_0^\varepsilon \to x$. Then $(X_t^\varepsilon, t \in [0, T])$ satisfies the LDP on $C([0, T], \mathbb{R})$ with rate function $S_{[0,T]}$ defined as follows.

$$S_{[0,T]}(\varphi) = \frac{1}{2} \int_{[0,T]} |\dot{\varphi}_t - (-U'(\varphi_t))|^2 dt$$

if $\varphi_0 = x$ and $t \mapsto \varphi_t$ is absolutely continuous; $S_{[0,T]}(\varphi) = +\infty$ otherwise.

Mean-field example – Large deviations in path-space

- $(\mu_N(t), t \in [0, T])$ is a $D([0, T], M_1(\mathcal{Z}))$-valued random variable.

Mean-field example – Large deviations in path-space

- $(\mu_N(t), t \in [0, T])$ is a $D([0, T], M_1(\mathcal{Z}))$-valued random variable.

Theorem

Let $\mu_N(0) \to \nu$ *weakly.*

Mean-field example – Large deviations in path-space

- $(\mu_N(t), t \in [0, T])$ is a $D([0, T], M_1(\mathcal{Z}))$-valued random variable.

Theorem

Let $\mu_N(0) \to \nu$ weakly. Then $\{\mu_N\}$ satisfies the LDP on $D([0, T], M_1(\mathcal{Z}))$ with rate function $S_{[0,T]}(\cdot)$ defined as follows.

Mean-field example – Large deviations in path-space

- $(\mu_N(t), t \in [0,T])$ is a $D([0,T], M_1(\mathcal{Z}))$-valued random variable.

Theorem

Let $\mu_N(0) \to \nu$ weakly. Then $\{\mu_N\}$ satisfies the LDP on $D([0,T], M_1(\mathcal{Z}))$ with rate function $S_{[0,T]}(\cdot)$ defined as follows. If $\mu(0) = \nu$ and $[0,T] \ni t \mapsto \mu_t \in M_1(\mathcal{Z})$ is absolutely continuous,

$$S_{[0,T]}(\mu) = \sup_{\alpha \in C([0,T], \mathbb{R}^{|\mathcal{Z}|})} \int_{[0,T]} \Bigg\{ \langle \alpha_t, \dot{\mu}_t - \Lambda^*_{\mu_t}\mu_t \rangle - \sum_{(z,z') \in \mathcal{E}} \tau(\alpha_t(z') - \alpha_t(z)) \lambda_{z,z'}(\mu_t)\mu_t(z) \Bigg\} dt,$$

$S_{[0,T]}(\mu) = +\infty$ otherwise.

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

Some remarks

- The log-moment generating function of the underlying noise appear in the rate function
  - $x \mapsto x^2/2$ for Gaussian.
  - $x \mapsto e^x - x - 1$ for Poisson.

Some remarks

- The log-moment generating function of the underlying noise appear in the rate function
  - $x \mapsto x^2/2$ for Gaussian.
  - $x \mapsto e^x - x - 1$ for Poisson.
- Rate function is the cost to produce a given trajectory
  - $X_t = \int_{[0,T]} -U'(X_t)dt + \varphi_t$ in the Gaussian case.
  - $\mu_t = \int_{[0,T]} \Lambda^*_{\mu_t} \mu_t dt + h_t$ in the mean-field case.

Some remarks

- The log-moment generating function of the underlying noise appear in the rate function
  - $x \mapsto x^2/2$ for Gaussian.
  - $x \mapsto e^x - x - 1$ for Poisson.
- Rate function is the cost to produce a given trajectory
  - $X_t = \int_{[0,T]} -U'(X_t)dt + \varphi_t$ in the Gaussian case.
  - $\mu_t = \int_{[0,T]} \Lambda^*_{\mu_t}\mu_t dt + h_t$ in the mean-field case.
- The variational form in the mean-field rate function is due to the non-uniqueness of h.

Some remarks

- The log-moment generating function of the underlying noise appear in the rate function
  - $x \mapsto x^2/2$ for Gaussian.
  - $x \mapsto e^x - x - 1$ for Poisson.
- Rate function is the cost to produce a given trajectory
  - $X_t = \int_{[0,T]} -U'(X_t)dt + \varphi_t$ in the Gaussian case.
  - $\mu_t = \int_{[0,T]} \Lambda^*_{\mu_t}\mu_t dt + h_t$ in the mean-field case.
- The variational form in the mean-field rate function is due to the non-uniqueness of h.
- Many proof techniques in the mean-field case
  - Sanov's theorem + Girsanov transformation (Dawson and Gärtner, Léonard).
  - Characterise the rate function using exponential martingales (Puhalskii, Liptser).
  - Weak convergence approach using variational formulae for Poisson random measure (Boué and Dupuis, Budhiraja).
  - Using semigroup convergence (Feng and Kurtz).

Part 3: Large deviations in the stationary regime

The unique equilibrium case

- Focus on the mean-field example from now on.

The unique equilibrium case

- Focus on the mean-field example from now on.
- Suppose that the McKean-Vlasov equation

$$\dot{\mu}_t = \Lambda^*_{\mu_t} \mu_t$$

has a unique globally asymptotically stable equilibrium ξ^*.

The unique equilibrium case

- Focus on the mean-field example from now on.
- Suppose that the McKean-Vlasov equation

$$\dot{\mu}_t = \Lambda^*_{\mu_t}\mu_t$$

  has a unique globally asymptotically stable equilibrium ξ^*.
- Is $\lim_{N\to\infty}\lim_{t\to\infty}\mu_N(t) \stackrel{?}{=} \lim_{t\to\infty}\lim_{N\to\infty}\mu_N(t)$
  - Can we study the stationary behaviour of the process μ_N by studying the large time behaviour of the McKean-Vlasov equation?

The unique equilibrium case

- Focus on the mean-field example from now on.
- Suppose that the McKean-Vlasov equation

$$\dot{\mu}_t = \Lambda^*_{\mu_t} \mu_t$$

  has a unique globally asymptotically stable equilibrium ξ^*.
- Is $\lim_{N\to\infty} \lim_{t\to\infty} \mu_N(t) \stackrel{?}{=} \lim_{t\to\infty} \lim_{N\to\infty} \mu_N(t)$
  - Can we study the stationary behaviour of the process μ_N by studying the large time behaviour of the McKean-Vlasov equation?
- We know that the RHS is ξ^*.

The unique equilibrium case

- Focus on the mean-field example from now on.
- Suppose that the McKean-Vlasov equation

$$\dot{\mu}_t = \Lambda^*_{\mu_t}\mu_t$$

  has a unique globally asymptotically stable equilibrium ξ^*.
- Is $\lim_{N\to\infty}\lim_{t\to\infty}\mu_N(t) \stackrel{?}{=} \lim_{t\to\infty}\lim_{N\to\infty}\mu_N(t)$
  - Can we study the stationary behaviour of the process μ_N by studying the large time behaviour of the McKean-Vlasov equation?
- We know that the RHS is ξ^*.
- Turns out that

$$\wp_N \to \delta_{\xi^*}$$

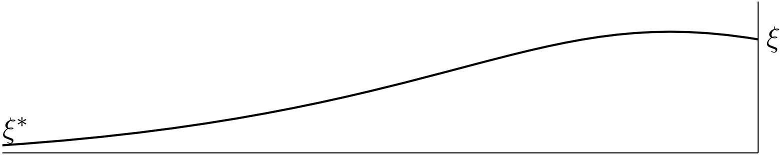
  weakly as $N \to \infty$.

The unique equilibrium case: Large deviations of $\wp_N$

- At stationarity, we expect $\mu_N(t)$ to be close to δ_{ξ^*}.

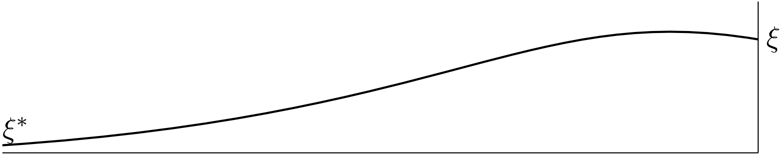
The unique equilibrium case: Large deviations of $\wp_N$

- At stationarity, we expect $\mu_N(t)$ to be close to δ_{ξ^*}.
- If $\mu_N(t)$ is near $\xi \neq \xi^*$, there is a cost of transport from ξ^* to ξ.



The unique equilibrium case: Large deviations of $\wp_N$

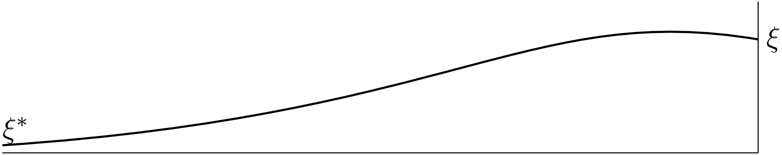
- At stationarity, we expect $\mu_N(t)$ to be close to δ_{ξ^*}.
- If $\mu_N(t)$ is near $\xi \neq \xi^*$, there is a cost of transport from ξ^* to ξ.



- Guess the cost:
  $V(\xi^*, \xi) = \inf\{S_{[-T,0]}(\mu), \mu_{-T} = \xi^*, \mu_0 = \xi, T > 0\}$.
- $V(\xi^*, \cdot)$ governs the LDP for $\{\wp_N\}$.

The unique equilibrium case: Large deviations of $\wp_N$

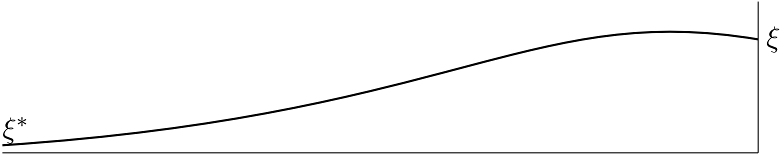
- At stationarity, we expect $\mu_N(t)$ to be close to δ_{ξ^*}.
- If $\mu_N(t)$ is near $\xi \neq \xi^*$, there is a cost of transport from ξ^* to ξ.



- Guess the cost:
  $V(\xi^*, \xi) = \inf\{S_{[-T,0]}(\mu), \mu_{-T} = \xi^*, \mu_0 = \xi, T > 0\}$.
- $V(\xi^*, \cdot)$ governs the LDP for $\{\wp_N\}$.
- How to show this?
  - Terminal condition is fixed but "initial condition" varies. Some uniform versions of the LDP of μ_N is needed.

The unique equilibrium case: Large deviations of $\wp_N$

- At stationarity, we expect $\mu_N(t)$ to be close to δ_{ξ^*}.
- If $\mu_N(t)$ is near $\xi \neq \xi^*$, there is a cost of transport from ξ^* to ξ.



- Guess the cost:
  $V(\xi^*, \xi) = \inf\{S_{[-T,0]}(\mu), \mu_{-T} = \xi^*, \mu_0 = \xi, T > 0\}$.
- $V(\xi^*, \cdot)$ governs the LDP for $\{\wp_N\}$.
- How to show this?
  - Terminal condition is fixed but "initial condition" varies. Some uniform versions of the LDP of μ_N is needed.
  - Natural to ask if
    $\inf\{S_{[-T,0]}(\mu), \mu_{-T} = \xi^*, \mu_0 = \xi, T > 0\} \stackrel{?}{=}$
    $\inf\{S_{(-\infty,0]}(\mu), \mu_t \to \xi^*, \mu_0 = \xi, T > 0\}$
    $\geq$ is clear. $\leq$ needs a "small-cost connection" from ξ^* to a nearby point.

Large deviations of $\mu_N(T)$

- The mapping $\pi_T : D([0, T], M_1(\mathcal{Z})) \to M_1(\mathcal{Z})$ is continuous. Use the contraction principle.

Theorem

Let $\mu_N(0) \to \nu$ *in weakly.*

Large deviations of $\mu_N(T)$

- The mapping $\pi_T : D([0,T], M_1(\mathcal{Z})) \to M_1(\mathcal{Z})$ is continuous. Use the contraction principle.

Theorem

Let $\mu_N(0) \to \nu$ in weakly. Then $\{\mu_N(T)\}_{T\geq 1}$ satisfies the LDP on $M_1(\mathcal{Z})$ with rate function

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

Large deviations of $\mu_N(T)$

- The mapping $\pi_T : D([0,T], M_1(\mathcal{Z})) \to M_1(\mathcal{Z})$ is continuous. Use the contraction principle.

Theorem

Let $\mu_N(0) \to \nu$ in weakly. Then $\{\mu_N(T)\}_{T \geq 1}$ satisfies the LDP on $M_1(\mathcal{Z})$ with rate function

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

- Note that $V(\nu, \xi) = \inf\{S_T(\xi|\nu), T > 0\}$.

"Cost" to go from ν to ξ in arbitrary time.

Large deviations of $\mu_N(T)$

- The mapping $\pi_T : D([0,T], M_1(\mathcal{Z})) \to M_1(\mathcal{Z})$ is continuous. Use the contraction principle.

Theorem

Let $\mu_N(0) \to \nu$ in weakly. Then $\{\mu_N(T)\}_{T \geq 1}$ satisfies the LDP on $M_1(\mathcal{Z})$ with rate function

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

- Note that $V(\nu, \xi) = \inf\{S_T(\xi|\nu), T > 0\}$.

"Cost" to go from ν to ξ in arbitrary time.

- If $\mathcal{Z}$ is a finite set, we can show that $S_T(\cdot|\cdot)$ is continuous (hence uniformly continuous). "Small-cost connection" follows.

Uniform large deviations

- A stronger version of the Freidlin-Wentzell definition.

Uniform large deviations

- A stronger version of the Freidlin-Wentzell definition.

Definition

$\{\mu_N\}$ is said to satisfy the uniform LDP over a class of subsets $\mathcal{A} \subset M_1(\mathcal{Z})$ if

Uniform large deviations

- A stronger version of the Freidlin-Wentzell definition.

Definition

$\{\mu_N\}$ is said to satisfy the uniform LDP over a class of subsets $\mathcal{A} \subset M_1(\mathcal{Z})$ if

- For each $K \subset S$ compact and $s \geq 0$, $\mathcal{K} = \bigcup_{\nu \in K} \Phi_\nu(s)$ is a compact subset of $D([0, T], M_1(\mathcal{Z}))$.

Uniform large deviations

- A stronger version of the Freidlin-Wentzell definition.

Definition

$\{\mu_N\}$ is said to satisfy the uniform LDP over a class of subsets $\mathcal{A} \subset M_1(\mathcal{Z})$ if

- For each $K \subset S$ compact and $s \geq 0$, $\mathcal{K} = \bigcup_{\nu \in K} \Phi_\nu(s)$ is a compact subset of $D([0, T], M_1(\mathcal{Z}))$.
- for any $\gamma > 0, \delta > 0, s > 0$ and $A \in \mathcal{A}$ there exists $N_0 \geq 1$ such that

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

for all $\nu \in A$, $\varphi \in \Phi_y(s)$ and $N \geq N_0$;

Uniform large deviations

- A stronger version of the Freidlin-Wentzell definition.

Definition

$\{\mu_N\}$ is said to satisfy the uniform LDP over a class of subsets $\mathcal{A} \subset M_1(\mathcal{Z})$ if

- For each $K \subset S$ compact and $s \geq 0$, $\mathcal{K} = \bigcup_{\nu \in K} \Phi_\nu(s)$ is a compact subset of $D([0, T], M_1(\mathcal{Z}))$.
- for any $\gamma > 0, \delta > 0, s > 0$ and $A \in \mathcal{A}$ there exists $N_0 \geq 1$ such that

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

  for all $\nu \in A$, $\varphi \in \Phi_y(s)$ 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

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

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

- In many cases, one can essentially "transfer" the sample-path LDP to this uniform version (usually, $\mathcal{A}$ is the power set).

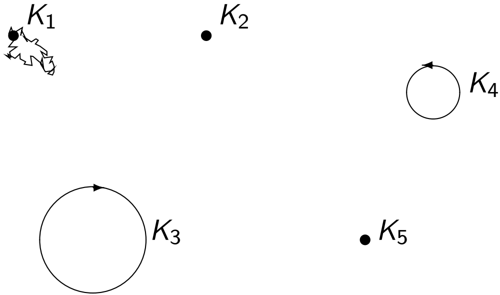
Uniform LDP – Examples

- Freidlin and Wentzell (1984): Finite-dimensional diffusions.
- Sowers (1992): A reaction diffusion equation (RDE).
- Cerrai and Röckner (2004): A multidimensional RDE.
- Borkar and Sundaresan (2012): Our mean-field example.
- Budhiraja et al (2018): A class of Banach-space valued SDEs.

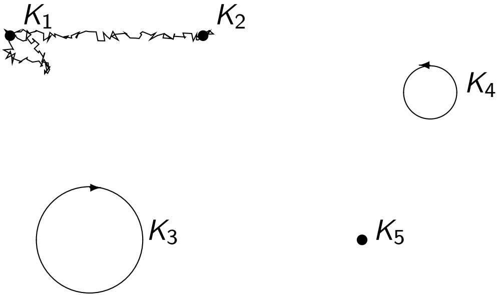
Uniform LDP – Examples

- Freidlin and Wentzell (1984): Finite-dimensional diffusions.
- Sowers (1992): A reaction diffusion equation (RDE).
- Cerrai and Röckner (2004): A multidimensional RDE.
- Borkar and Sundaresan (2012): Our mean-field example.
- Budhiraja et al (2018): A class of Banach-space valued SDEs.
- Unique equilibrium case: Uniform LDP on compact sets + "small-cost" connection.

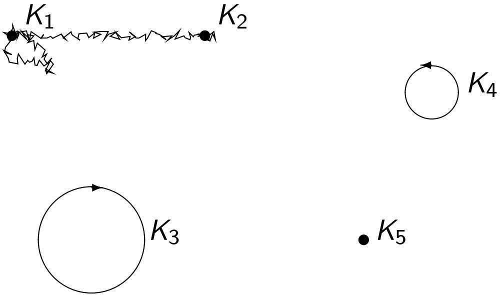
Multiple equilibria



Multiple equilibria



Multiple equilibria



- τ_n: hitting time of μ_N in a given neighbourhood of K_i's.
- Hitting time chain: $Z_n^N = \mu_N(\tau_n)$, $n \geq 1$.

Multiple equilibria

- First, consider a hitting-time Markov chain Z^N.
- Z^N is a finite state-space chain.

Multiple equilibria

- First, consider a hitting-time Markov chain Z^N.
- Z^N is a finite state-space chain.
- Transition probability of Z^N is of the form $P(K_i, K_j) \sim \exp\{-N\tilde{V}(K_i, K_j)\}$.
- Invariant measure for Z^N can be obtained (Markov chain tree theorem).

Multiple equilibria

- First, consider a hitting-time Markov chain Z^N.
- Z^N is a finite state-space chain.
- Transition probability of Z^N is of the form $P(K_i, K_j) \sim \exp\{-N\tilde{V}(K_i, K_j)\}$.
- Invariant measure for Z^N can be obtained (Markov chain tree theorem).
- Khasminskii formula: Reconstruct the invariant measure of μ_N from that of Z^N.

Multiple equilibria

- First, consider a hitting-time Markov chain Z^N.
- Z^N is a finite state-space chain.
- Transition probability of Z^N is of the form $P(K_i, K_j) \sim \exp\{-N\tilde{V}(K_i, K_j)\}$.
- Invariant measure for Z^N can be obtained (Markov chain tree theorem).
- Khasminskii formula: Reconstruct the invariant measure of μ_N from that of Z^N.
- $V(\xi) = \min_{1\leq i\leq l}(W(K_i) + V(K_i, \xi)) - \min_{1\leq i\leq l} W_i$. W_i describes the invariant measure of Z^N.

Multiple equilibria

- First, consider a hitting-time Markov chain Z^N.
- Z^N is a finite state-space chain.
- Transition probability of Z^N is of the form $P(K_i, K_j) \sim \exp\{-N\tilde{V}(K_i, K_j)\}$.
- Invariant measure for Z^N can be obtained (Markov chain tree theorem).
- Khasminskii formula: Reconstruct the invariant measure of μ_N from that of Z^N.
- $V(\xi) = \min_{1 \le i \le l}(W(K_i) + V(K_i, \xi)) - \min_{1 \le i \le l} W_i$. W_i describes the invariant measure of Z^N.
- Key properties used:
  - Strong Markov property of μ_N.
  - Uniform LDP for μ_N on open sets of $M_1(\mathcal{Z})$.
  - Continuity of the quasipotential V.

Some technicalities in infinite dimensions

- Examples: Stochastic PDEs, mean-field model on $\mathbb{Z}_+$.

Some technicalities in infinite dimensions

- Examples: Stochastic PDEs, mean-field model on $\mathbb{Z}_+$.
- Need to show the existence and uniqueness of the invariant measure $\wp_N$ under suitable assumptions.

Some technicalities in infinite dimensions

- Examples: Stochastic PDEs, mean-field model on $\mathbb{Z}_+$.
- Need to show the existence and uniqueness of the invariant measure $\wp_N$ under suitable assumptions.
- $M_1(\mathbb{Z}_+)$ is not locally compact. Since $S_T(\cdot|\nu)$ has compact level sets, it cannot be continuous.
  - Example: $D(\cdot\|\nu)$ is not continuous on $M_1(\mathcal{Z})$. It has compact level sets (Donsker-Varadhan variational formula).

Some technicalities in infinite dimensions

- Examples: Stochastic PDEs, mean-field model on $\mathbb{Z}_+$.
- Need to show the existence and uniqueness of the invariant measure $\wp_N$ under suitable assumptions.
- $M_1(\mathbb{Z}_+)$ is not locally compact. Since $S_T(\cdot|\nu)$ has compact level sets, it cannot be continuous.
  - Example: $D(\cdot\|\nu)$ is not continuous on $M_1(\mathcal{Z})$. It has compact level sets (Donsker-Varadhan variational formula).
- V cannot be continuous. Some regularity of V needs to be shown.

Some technicalities in infinite dimensions

- Examples: Stochastic PDEs, mean-field model on $\mathbb{Z}_+$.
- Need to show the existence and uniqueness of the invariant measure $\wp_N$ under suitable assumptions.
- $M_1(\mathbb{Z}_+)$ is not locally compact. Since $S_T(\cdot|\nu)$ has compact level sets, it cannot be continuous.
  - Example: $D(\cdot\|\nu)$ is not continuous on $M_1(\mathcal{Z})$. It has compact level sets (Donsker-Varadhan variational formula).
- V cannot be continuous. Some regularity of V needs to be shown.
  - E.g. $V(\xi^*, \cdot)$ is finite on measures with finite second moment, and continuous on measures whose $(2+\varepsilon)$ moment is bounded.

Some technicalities in infinite dimensions

- Examples: Stochastic PDEs, mean-field model on $\mathbb{Z}_+$.
- Need to show the existence and uniqueness of the invariant measure $\wp_N$ under suitable assumptions.
- $M_1(\mathbb{Z}_+)$ is not locally compact. Since $S_T(\cdot|\nu)$ has compact level sets, it cannot be continuous.
  - Example: $D(\cdot\|\nu)$ is not continuous on $M_1(\mathcal{Z})$. It has compact level sets (Donsker-Varadhan variational formula).
- V cannot be continuous. Some regularity of V needs to be shown.
  - E.g. $V(\xi^*, \cdot)$ is finite on measures with finite second moment, and continuous on measures whose $(2+\varepsilon)$ moment is bounded.
- The uniform LDP needs be shown.
  - In the multiple equilibria case, existence of the invariant measure for Z^N is nontrivial.

Some technicalities in infinite dimensions

- Examples: Stochastic PDEs, mean-field model on $\mathbb{Z}_+$.
- Need to show the existence and uniqueness of the invariant measure $\wp_N$ under suitable assumptions.
- $M_1(\mathbb{Z}_+)$ is not locally compact. Since $S_T(\cdot|\nu)$ has compact level sets, it cannot be continuous.
  - Example: $D(\cdot\|\nu)$ is not continuous on $M_1(\mathcal{Z})$. It has compact level sets (Donsker-Varadhan variational formula).
- V cannot be continuous. Some regularity of V needs to be shown.
  - E.g. $V(\xi^*, \cdot)$ is finite on measures with finite second moment, and continuous on measures whose $(2+\varepsilon)$ moment is bounded.
- The uniform LDP needs be shown.
  - In the multiple equilibria case, existence of the invariant measure for Z^N is nontrivial.
  - Khasminskii reconstruction needs to be extended.

Some recent progress

- Martirosyan (2017): Considers a situation where an invariant measure for Z^N may not exist. Works with "generalised stationary distributions".
- Puhalskii (2020): Provides some general conditions under which LDP for the invariant measure can be shown from the process-level LDP.
- Budhiraja et al (2018): Derives uniform LDP for a Banach space valued SDE using the weak convergence approach.
- Salins and Spiliopoulos (2018): Shows some continuity properties of the quasipotential for an infinite dimensional SDE.

Some open questions

- Uniform LDP in infinite dimensions for the mean-field setting
  - Not aware of such a result even in the non-interacting case.
  - However, uniform LDP on compact sets can be established (this is sufficient for the problem of invariant measure).
- A general result from which LDP for the invariant measure can be derived from process-level LDP
  - Puhalskii (2020) answers this partially, but this is not applicable in infinite dimensional settings.
  - Borkar and Sundaresan (2012) does this for the mean-field example using a control theoretic approach.

Conclusion

- Introduction to the metastability phenomenon and the theory of large deviations.

Conclusion

- Introduction to the metastability phenomenon and the theory of large deviations.
- Broadly, two techniques to go from the sample path LDP to the invariant measure LDP:
    - Unique equilibrium case: use a small-cost connection. Uniform LDP on compact sets suffices.
    - Multiple equilibria: Hitting time chain + Khasminskii reconstruction. Needs the full uniform LDP.

Conclusion

- Introduction to the metastability phenomenon and the theory of large deviations.
- Broadly, two techniques to go from the sample path LDP to the invariant measure LDP:
  - Unique equilibrium case: use a small-cost connection. Uniform LDP on compact sets suffices.
  - Multiple equilibria: Hitting time chain + Khasminskii reconstruction. Needs the full uniform LDP.
- References: Freidlin and Wentzell (1984), Dawson and Gärtner (1987), Sowers (1990), Léonard (1995), Cerrai and Röckner (2004), Bordenave et al (2009), Borkar and Sundaresan (2012), Martirosyan (2017), Budhiraja et al (2018), Salins and Spiliopoulos (2018), Puhalskii (2020).

Thank you