Mean-field limits in communication networks, and a closer look at the fixed-point method

Rajesh Sundaresan

Indian Institute of Science

27-29 July 2015

Outline

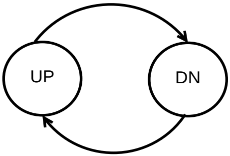
1. Model description, finite and large-time mean-field limits, and the fixed-point method.

2. Large deviation from the mean-field limit for finite time durations.

3. Large deviation from the mean-field limit for the stationary measure.

A mean-field model of a spin system

- Interacting system with N particles
- Each particle's state space: $\mathcal{Z} = \{\text{UP}, \text{DN}\}$
- Transitions:



- Dynamics depends on the "mean field". Global interaction.
  $\mu_N(t) =$ Fraction of particles having UP spin
- Transition from i to j at rate $\lambda_{ij}(\mu_N(t))$

Reversible versus nonreversible dynamics

- (Reversible) Gibbsian system
  - Example: Heat bath dynamics
  - $E(\mu_N)$: Energy of a configuration $x = (x_1, \ldots, x_N)$ with mean μ_N
  - An i to j transition takes μ_N to $\mu - \frac{1}{N}\delta_i + \frac{1}{N}\delta_j$

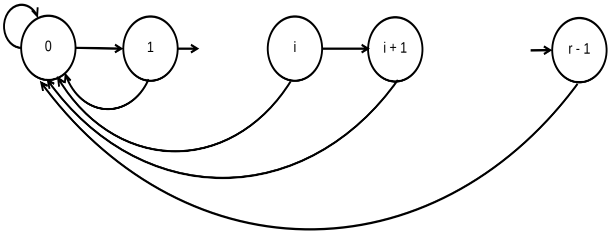
$$\lambda_{ij}(\mu_N) = \frac{e^{-NE(\mu_N)}}{e^{-NE(\mu_N - \frac{1}{N}\delta_i + \frac{1}{N}\delta_j)} + e^{-NE(\mu_N)}}$$

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

- Weak interaction

Wireless Local Area Network (WLAN) interactions

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



- Interpretation
  - 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

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 $\lambda_{i,j}(\mu_N(t))$

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 $c = (c_0, c_1, c_2)$.
- For μ, the empirical measure of a configuration, the rate matrix is

$$\Lambda(\mu) = \left[\begin{array}{ccc} -(\cdot) & c_0(1 - e^{-\langle c,\mu \rangle}) & 0 \\ c_1 e^{-\langle c,\mu \rangle} & -(\cdot) & c_1(1 - e^{-\langle c,\mu \rangle}) \\ c_2 & 0 & -(\cdot) \end{array} \right].$$

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

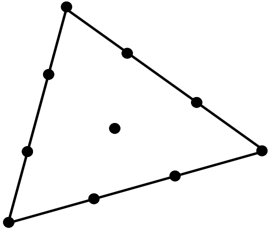
Engineering: Going the full cycle

- Design the protocol. This fixes the interaction and the dynamics.
  - Allow ourselves flexibility. Enough parameters to tune.
    Here, aggressiveness c.
- Analysis/Simulation: Study phenomena as a function of the parameters.
- Choose parameters. Choice guided by studies in the previous step.
- At a slower time-scale, change the protocol.
  Flaws in protocol. Or newer requirements.
  Capture model (Neelesh Mehta and his team).
  IEEE 1901 (P. Thiran and his team).

These talks: analysis.
Example happy situation where analysis justified a simplifying approximation, explained phenomena observed in simulations and in practice. Also, general enough to handle capture model, IEEE 1901 model etc.

The Markov processes, big and small

- $(X_n^{(N)}(\cdot),\ 1 \leq n \leq N)$ is Markov
- State space grows exponentially with N: size r^N
- Study $\mu_N(\cdot)$ instead, also a Markov process
  Its state space size is at most $(N+1)^r$, and is a subset of $\mathcal{M}_1(\mathcal{Z})$
  Then try to draw conclusions on the original process.
- State space of $\mu_N(\cdot)$



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

- A Markov process with state space being the set of empirical measures of N particles.
- This is a measure-valued flow across time.
- The transition from μ to $\mu + \frac{1}{N}e_j - \frac{1}{N}e_i$ occurs with rate $N\mu(i)\lambda_{i,j}(\xi)$.
- 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 $\Lambda(\cdot) = [\ \lambda_{i,j}(\cdot)\ ]$. Then

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

- Interpretation: The rate of change in the kth component is

$$\sum_{i:i \neq k} \xi_i \lambda_{i,k}(\xi) - \xi_k \sum_{i:i \neq k} \lambda_{k,i}(\xi)$$

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

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

- Nonlinear ODE. A transport equation. Lives in $\mathcal{M}_1(\mathcal{Z})$.

Assumptions

- The graph with vertex set $\mathcal{Z}$ and edge set $\mathcal{E}$ is irreducible
  Holds in our WLAN example

- There exist positive constants $c > 0$ and $C < +\infty$ such that, for every $(i, j) \in \mathcal{E}$, we have

$$c \leq \lambda_{i,j}(\cdot) \leq C$$

- The mapping $\mu \mapsto \lambda_{i,j}(\mu)$ is Lipschitz continuous over $\mathcal{M}_1(\mathcal{Z})$

The notion of convergence

- $\mu_N(\cdot)$ takes values in $D([0, T], \mathcal{M}_1(\mathcal{Z}))$, right-continuous with left limits, measure-valued paths.

- Equip this space with the metric

$$\rho_T(\eta(\cdot), \zeta(\cdot)) = \sup_{t \in [0,T]} ||\eta(t) - \zeta(t)||_1$$

  where $|| \cdot ||_1$ is the L^1 metric.

- Convergence is uniform over $[0, T]$.

Kurtz's theorem: a formal statement

Let $\mu(\cdot)$ be the solution to the McKean-Vlasov dynamics with initial condition $\mu(0) = \nu$.

Theorem

Let $\mu_N(0) \xrightarrow{p} \nu$, where ν is deterministic. Then, for each $T > 0$, $\mu_N(\cdot) \xrightarrow{p} \mu(\cdot)$.

Remarks:

- The McKean-Vlasov ODE must be well-posed. Lipschitz suffices.
- $\mu_N(0) \xrightarrow{p} \nu$: Probability of being outside a ball around ν vanishes.
- $\mu_N(\cdot) \xrightarrow{p} \mu(\cdot)$: For any finite duration, probability of being outside a tube around $\mu(\cdot)$ vanishes.

Proof methods

- Get estimates on $\rho_T(\mu_N, \mu)$ via Gronwall bound and show that the probability that it exceeds ε vanishes with N.

- Or show that the infinitesimal generator $\mathscr{L}_N$ for the Markov process $\mu_N()$ converges to a first order differential operator.

  For any bounded and continuous $\Phi : \mathcal{M}_1(\mathcal{Z}) \to \mathbb{R}$, the function $\mathscr{L}_N \Phi$ is the conditional expected drift starting from the argument ξ:

$$
\begin{aligned}
\mathscr{L}_N \Phi(\xi) &= \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}\left[\Phi(\mu_N(t+h)) - \Phi(\mu_N(t)) \mid \mu_N(t) = \xi\right] \\
&= \sum_{(i,j): j \neq i} N\xi(i)\lambda_{i,j}(\xi) \left[\Phi\left(\xi + \frac{1}{N}e_j - \frac{1}{N}e_i\right) - \Phi(\xi)\right] \\
&= \langle \nabla \Phi(\xi), \Lambda(\xi)^* \xi \rangle + O\left(\frac{1}{N}\right)
\end{aligned}
$$

via Taylor if Φ has bounded second order derivatives.

Back to the individual particles

- Let $\mu(\cdot)$ be the solution to the McKean-Vlasov dynamics
- Tag a particle.
  - Its evolution influenced by the mean-field $\mu_N(\cdot)$.
  - But the mean-field $\mu_N(\cdot)$ converges to a deterministic limit.
  - Asymptotically then, the particle executes a Markov process with time-dependent transition rates $\lambda_{i,j}(\mu(t))$
  - Can formalise this notion.
  - $\mu(t)$ is the distribution for the state of the tagged particle at time t.

Joint evolution of tagged particles

- Tag k particles.
  - Exchangeable $X^N(0)$. Take the limit as $N \to \infty$.
  - De Finetti's theorem: An infinite exchangeable process is a mixture of iids.
    The "driving" distribution is the distribution of $\lim_N \mu_N(0)$.
  - If exchangeable, and $\mu_N(0) \xrightarrow{p} \nu$ (deterministic), then the particle states are asymptotically independent at time 0.
  - Chaoticity or "Boltzmann property" (M. Kac 1956).
  - If Boltzmann property holds at time 0, Boltzmann property holds at any time $t > 0$. (Kac 1956)
  - Initial chaos propagates over time.
- "Canonical ensemble" at time t.

Large time behaviour

- $\lim_{t\to\infty}[\lim_{N\to\infty}\mu_N(t)]$ reduces to a study of the McKean-Vlasov ODE for large time.

- $\lim_{N\to\infty}[\lim_{t\to\infty}\mu_N(t)]$?
  - For fixed N, the time limit continues to be random. Let $\wp^{(N)} = Law(\mu_N(\infty))$.
  - The stationary distribution exists and is unique by our assumptions.

- What is $\lim_{N\to\infty}\wp^{(N)}$?
  Does the first limit say something about this?

Some inescapable terminology on dynamical systems

- ODE: $\dot{\mu}(t) = F(\mu(t))$ for $t \geq 0$ with initial condition $\mu(0) = \nu$.
- Stationary point: Solutions to $F(\xi) = 0$.
- ω-limit set $\Omega(\nu)$: All limit points of $\mu(\cdot)$ when $\mu(0) = \nu$.
- Recurrent point: A ν such that $\nu \in \Omega(\nu)$.
- Birkhoff centre $\mathcal{B}$ = set of all recurrent points.
- Example: $\dot{\mu}(t) = A\mu(t)$, with A nonsingular, $\mu(t) \in \mathbb{R}^2$.
  - Stationary points = $\{0\}$.
  - Case when all Re $\lambda_i < 0$: $\mathcal{B} = \{0\}$.
  - Case when all Re $\lambda_i = 0$: $\mathcal{B} = \{0\} \cup$ all circles $= \mathbb{R}^2$.
- A stationary point ξ_0 is globally asymptotically stable if (among other things) $\mu(t) \to \xi_0$ for all initial conditions $\mu(0)$.
  In particular, $\mathcal{B} = \{\xi_0\}$

The limiting behaviour of the stationary distribution

Theorem

- *The support of any limit of* $(\wp^{(N)}, N \geq 1)$ *is a compact subset of the Birkhoff centre* $\mathcal{B}$.
- *If* ξ_0 *is a stationary point that is globally asymptotically stable for the McKean-Vlasov dynamics, then* $\wp^{(N)} \to \delta_{\xi_0}$*, that is,* $\mu_N(\infty) \to \xi_0$ *in distribution (and hence in probability).*
- *Decoupling: Tag* k *nodes* $n_1, n_2, \ldots, n_k$*. Then*

$$\left(X_{n_1}^{(N)}(\infty), X_{n_2}^{(N)}(\infty), \ldots, X_{n_k}^{(N)}(\infty)\right) \to \xi_0^{\otimes k} \text{ in distribution.}$$

We will not discuss the proof. But the first two are consequences of a more general result (to be covered in a later lecture).

Stationary points, fixed-points, and all that

- Stationary point of the dynamics: Solve for ξ in $\Lambda(\xi)^* \xi = 0$.

- Fixed-point analysis
  - Assume that a tagged particle has distribution ξ_0 in steady state. Assume symmetry – all particles have the same steady state distribution. This sets up the field $\Lambda(\xi_0)^*$ for the tagged particle. The field must be such that ξ_0 is fixed.
  - Solving for stationary points.
  - In some cases, can look for simpler interpretable macroscopic variables - attempt probabilities or collision probabilities. In the WLAN case, a fixed point equation in one variable (e.g., collision probability).

- Take ξ_0 as describing the steady state behaviour of the system.

Limitation of the fixed-point analysis

- There may be a unique stationary point, but it may not be globally asymptotically stable.
  - Benaim and Le Boudec have an example where stationary point is unique, but unstable. All trajectories converge to a limit cycle.
- If $c_i = c_0/2^i$ and $c_0 < \ln 2$, then there is a unique stationary point ξ_0 that is globally asymptotically stable.
- Three states with $c_0 = 0.5$, $c_1 = 0.3$ but $c_3 = 8$ (say). Three stationary points – two stable and one unstable.
- Since the finite N system can be viewed as the deterministic dynamical system with noise, the unstable points are not going to be in the support of any limit point of $\wp^{(N)}$.
- Question: Which of the multiple stable stationary points (or limit cycles) will best describe the large time behaviour of the system?

A Lyapunov function

- If the differential equation were linear, i.e., $\dot{\mu}(t) = \Lambda^* \mu(t)$ ...
  - The associated Markov process X^N does not have any interaction.
  - Let the stationary measure for one particle's evolution be π^*.
  - Relative entropy $I(\cdot|\pi^*)$ is a "Lyapunov function" for the dynamics.
  - $I(\mu(t)||\pi^*) \downarrow 0$ as time progresses, and $\mu(t) \to \pi^*$.

- Does a Lyapunov function exist for the nonlinear dynamical system?
- When is it global?
  When local, does it "select" the best stable equilibrium or equilibria?

Large deviations, Freidlin-Wentzell theory, quasipotential

- For the linear differential equation, and the associated (noninteracting) Markov process, let π^* denote the stationary distribution.

- The sequence of stationary distributions for $\mu_N(\cdot)$ satisfies a large deviation principle:

$$\Pr\{\mu_N(\infty) \in \text{ neighbourhood of } q\} \sim \exp\{-NI(q|\pi^*)\}$$

with rate function $I(\cdot|\pi^*)$

  - Independent samplings of π^* leading to the empirical measure $\mu_N(\infty)$. Apply Sanov's theorem.
  - This rate function serves as a Lyapunov function.

Large deviations, Freidlin-Wentzell theory, quasipotential

Do the same for the nonlinear differential equation, and its associated (weak interaction) Markov process.

Theorem (with V.S.Borkar)

Let the McKean-Vlasov dynamics have a globally asymptotically stable equilibrium π^*. Let $V(q)$ minimise the following actional functional.

$$V(q) = \inf \left\{ \int_0^T L(\phi(s), \dot{\phi}(s)) \, ds \mid \phi(0) = \pi^*, \phi(T) = q, T \in (0, \infty) \right\}.$$

Then the sequence of stationary distributions for $\mu_N(\cdot)$ satisfies an LDP with rate funtion V.

$L(\phi(t), \dot{\phi}(t)) = 0$ if $\dot{\phi}(t)$ obeys the nonlinear dynamics.
$V(q)$ plays the role of relative entropy.
(In particular, $V(\pi^*) = 0$.)

Can extend to metastable setting as well. See a later lecture.

Final remarks

- A local Lyapunov function in Gibbsian, locally Gibbsian systems. Budhiraja et al. (arXiv:1412.5555)
- Analogy: Second law of thermodynamics. Fixed-point analysis is an analysis of the system in equilibrium. Collision probability is a macroscopic variable. ξ_0 corresponds to the canonical ensemble.
- Metastability or not?
  - When no metastability, the decoupling approximation works.
  - Large deviation principle suggests exponentially fast concentration. Approximation is likely to be good.
- Design c to avoid metastability. Example: $\ln 2 > c_0 > 2c_1 > 4c_2$. Is this the best in terms of throughput without metastability?
- Current studies:
  - Best choice of c in the stable regime.
  - Newer protocols with particles of different classes (arising from different quality of service requirements).

References

1. Kac, M. (1956) "Foundations of kinetic theory", in '*Proc. of 3rd Berkeley Symp. on Math. Stat. and Prob.*' (J. Neyman, ed.), Univ. of California Press, 171–197.
2. McKean, H. P. (1966) "A class of Markov processes associated with nonlinear parabolic equations", *Proc. Nat. Acad. Sci.* **56**, 1907–1911.
3. Stolyar, A. L. (1989) "Asymptotic behavior of the stationary distribution for a closed queueing system", *Problemy Peredachi Informatsii*, **25**(4), 80–92.
4. Sznitman, A. S. (1991) "Topics in propogation of chaos", in '*Ecole d'été de Probabilités de Saint Flour (1989)*', Springer Lecture Notes in Mathematics No. **1464**, 166–251.
5. Anantharam, V. (1991) "A mean field limit for a lattice caricature of dynamic routing in circuit switched networks", *Annals of Appl. Prob.* **1**, 481–503.
6. Anantharam, V. and Benchekroun, M. (1993) "A technique for computing sojourn times in large networks of interacting queues", *Probability in the Engineering and Informational Sciences* **7**, 441–464.
7. Bianchi, G. (1998) "IEEE 802.11 - saturated throughput analysis", *IEEE Comm. Lett.* **12**, 318–320.
8. Graham, C. (2000) "Chaoticity on path space for a queueing network with selection of the shortest queue among several", *J. Appl. Probab.* **37**(1), 198–211.
9. Sharma, G., Ganesh, A. J. and Key, P. (2009) "Performance analysis of contention based medium access protocols", *IEEE Trans. Information Th.* **55**(4), 1665–1682. Also reported in *Proc. IEEE Infocom (2006)*.
10. Vvedenskaya, N. D. and Sukhov Yu. M. (2007), "A multi-user multiple-access system: stability and metastability", *Problemy Peredachi Informatsii* **43**, 105–111.
11. Ramaiyan, V., Kumar, A. and Altman, E. (2008) "Fixed point analysis of single cell IEEE 802.11e WLANs: uniqueness and multistability", *IEEE/ACM Trans. Network.* **16**, 1080–1093.
12. Benaim, M. and Le Boudec J.-Y. (2008) "A class of mean field interaction models for computer and communication systems", *Perform. Eval.* **65**, 823–838.
13. Bordenave, C., McDonald, D. and Proutière, A. (2010) "A particle system in interaction with a rapidly varying environment: mean field limits and applications", *Networks and Heterogeneous Media*, **5**(1).
14. Duffy, K. R. (2010) "Mean field Markov models of wireless local area networks", *Markov Processes and Related Fields* **16**(2), 295–328.

A primer on large deviations

- N iid tosses from a coin with bias $\Pr\{X_n = 1\} = \lambda \in (0, 1)$.
- Estimate of bias $\hat{\lambda}_N = \frac{1}{N}\sum_{n=1}^{N} X_n$.
- Weak LLN says $\hat{\lambda}_N \to \lambda$ (in probability)
- Assume $\tau > \lambda$. Chernoff bound says

  $$\Pr\left\{\frac{1}{N}\sum_{n=1}^{N} X_n \geq \tau\right\} \leq \exp\{-N \sup_{t\geq 0}\left[\tau \cdot t - \log \mathbb{E}e^{tX_1}\right]\} = \exp\{-NI(\tau||\lambda)\}$$

  where I is relative entropy

  $$I(\tau||\lambda) = \tau \log\left(\frac{\tau}{\lambda}\right) + (1-\tau)\log\left(\frac{1-\tau}{1-\lambda}\right).$$

  Cramer's theorem says this bound is tight in exponential scale, i.e.,

  $$\lim_{N\to\infty} \frac{1}{N} \log \Pr\{\cdots\} = -I(\tau||\lambda)$$

- Similarly for $\tau < \lambda$.

Deviations to more general sets

- If $A \subset [0,1]$ is some interval, then by chopping into small intervals, using continuity of I, we get

$$\lim_{N\to\infty} \frac{1}{N} \log \Pr\{\hat{\lambda}_N \in A\} = -\inf_{x\in A} I(x||\lambda).$$

- In general, $\inf\{I(x)|x \in A^\circ\}$ and $\inf\{I(x)|x \in \overline{A}\}$ may be different, and so

Large deviation principle (LDP)

- *Definition*: A sequence $(p^{(N)}, N \geq 1)$ of probability measures on a metric space $\mathcal{X}$ satisfies the LDP with speed N and good rate function $I(\cdot)$ if
  - For every open set G and closed set F of the metric space $\mathcal{X}$, we have

$$\liminf_{N \to +\infty} \frac{\log p^{(N)}(G)}{N} \geq -\inf_{x \in G} I(x)$$
$$\limsup_{N \to +\infty} \frac{\log p^{(N)}(F)}{N} \leq -\inf_{x \in F} I(x)$$

  - For each $a \in [0, \infty)$, the level sets $\{x : I(x) \leq a\}$ are compact

An aside: LDP for empirical measures

- Fact: Let $p^{(N)}$ be the law of $\frac{1}{N}\sum_{n=1}^{N} X_n \in [0,1]$. This sequence satisfies the LDP speed N and good rate function $I(\cdot||\lambda)$.

- $\frac{1}{N}\sum_{n=1}^{N} X_n$ may be viewed as an empirical measure $\frac{1}{N}\sum_{n=1}^{N} \delta_{X_n}$ on $\{0,1\}$.

- Restatement of fact:
  Let $p^{(N)}$ now be the law of $\frac{1}{N}\sum_{n=1}^{N} \delta_{X_n} \in \mathcal{M}_1(\{0,1\})$.
  This sequence satisfies the LDP with speed N and rate function

$$S(\{\tau, 1-\tau\}||\{\lambda, 1-\lambda\}) = I(\tau||\lambda).$$

Theorem (Sanov's theorem)

Let $\mathcal{X}$ be a Polish space and let $\mathcal{M}_1(\mathcal{X})$ be the space of probability measures on $\mathcal{X}$ equipped with the topology of weak convergence. Let $X_1, \cdots, X_N$ be sampled iid from P. Then the empirical measures satisfy the LDP on $\mathcal{M}_1(\mathcal{X})$ with good rate function $I(\cdot||P)$.

Back to μ_N ...

- Metric space $D([0,T],\mathcal{M}_1(\mathcal{Z}))$ (with metric ρ_T coming from sup-norm).
- $p^{(N)}_{\nu_N}$ is the law of $(\mu_N(t), t \in [0,T])$ starting at ν_N.
- Rate function will be a function of paths and will be denoted $S_{[0,T]}(\mu|\nu)$.
- There is dependence on the initial condition ν.

Finite duration LDP

Theorem

Suppose that the initial conditions $\nu_N \to \nu$*.*

Then the sequence $(p^{(N)}_{\nu_N}, N \geq 1)$ *satisfies the LDP on* $D([0,T], \mathcal{M}_1(\mathcal{Z}))$ *(with metric* ρ_T*) with speed* N *and a good rate function* $S_{[0,T]}(\mu|\nu)$*.*

If a path $\mu \in D([0,T], \mathcal{M}_1(\mathcal{Z}))$ *has* $S_{[0,T]}(\mu|\nu) < +\infty$*, then*

- *the time derivative* $\dot{\mu}$ *exists for almost all* $t \in [0,T]$*;*
- *there exist rates* $(l_{i,j}(t), t \in [0,T],\ (i,j) \in \mathcal{E})$ *such that*

$$\dot{\mu}(t) = L(t)^* \mu(t)$$

  where $L(t)$ *is the rate matrix associated with the time-varying rates* $(l_{i,j}(t), (i,j) \in \mathcal{E})$ *and* $L(t)^*$ *is its adjoint;*
- *the good rate function* $S_{[0,T]}(\mu|\nu)$ *is given by*

$$S_{[0,T]}(\mu|\nu) = \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} (\mu(t)(i))\lambda_{i,j}(\mu(t))\ \tau^* \left(\frac{l_{i,j}(t)}{\lambda_{i,j}(\mu(t))} - 1 \right) \Big] dt.$$

Proof outline

- Apply Sanov's theorem to noninteracting system on path space
- Relate the interacting system to the noninteracting system via Girsanov's formula
- Use the Laplace-Varadhan principle to extract a path space LDP for the interacting system
- Then use the contraction principle (from an LDP for the empirical measure in path space to an LDP for the law of $\mu_N(\cdot)$).

Corollary:
$p^{(N)}_{\nu_N} \to \delta_{\mu(\cdot)}$ weakly, where $\mu(\cdot)$ is the McKean-Vlasov solution

Proof steps in a little more detail

- Look at a larger object. Empirical measures on path space.

- Space of interest, measures on the space, topology
  - Given a particle's trajectory $x(\cdot)$, let $\phi(x) =$ number of jumps in $[0, T]$.
  - $\mathcal{X} = \{x(\cdot) \mid x \text{ has jumps in } \mathcal{E} \text{ and } \phi(x) < \infty\}$.
  - $d(x, y) = d_{Sko}(x, y) + |\phi(x) - \phi(y)|$. (Polish space)
  - Let f be continuous and define

$$||f||_\phi = \sup_{x \in \mathcal{X}} \frac{|f(x)|}{1 + \phi(x)}.$$

  - $C_\phi(\mathcal{X})$ is the set of continuous functions with finite norm.
  - $\mathcal{M}_{1,\phi}(\mathcal{X}) = \{Q \in \mathcal{M}_1(\mathcal{X}) \mid \int \phi \, dQ < \infty\}$.
  - Topology:
    $Q_N \to Q$ if and only if $\int f \, dQ_N \to \int f \, dQ$ for all $f \in C_\phi(\mathcal{X})$.
  - σ-field: cylinder σ-field on $\mathcal{M}_{1,\phi}(\mathcal{X})$.

The probability measures with and without interaction

- $\bar{P}_z$: Law of the Markov process where all allowed transition rates are 1, and initial condition is z.
- $P_z(\mu)$: Law of the Markov process where rate matrix at time t is $\Lambda(\mu(t))$, and initial condition is z.

- $\bar{\mathbb{P}}^{(N)}_{z^N}$: Law of the N particle evolutions without interaction with initial condition z^N.
- $\mathbb{P}^{(N)}_{z^N}$: Law of the N particle evolutions with interaction with initial condition z^N.
- $x^N(\cdot)$: description of evolution of all N particles, with identities preserved.
- $x^N(\cdot) \overset{G_N}{\mapsto} Q_N = \frac{1}{N}\sum_{n=1}^{N} \delta_{x_n(\cdot)}$: empirical measure.
- $Q_N \overset{\pi}{\mapsto} \mu_N$: from empirical measure to measure-valued process.

Girsanov's formula

- Using the independent increments property and the dependence of transition rates only on the mean-field, we can get a Girsanov formula:
$$\frac{d\mathbb{P}^{(N)}_{z^N}}{d\bar{\mathbb{P}}^{(N)}_{z^N}}(x^N) = e^{Nh(G_N(x^N))} = e^{Nh(Q_N)}.$$

- Let $P^{(N)}_{\nu_N}$ and $\bar{P}^{(N)}_{\nu_N}$ be the push forwards of the interacting and noninteracting distributions under $x^N \mapsto Q_N$. Then

$$\frac{dP^{(N)}_{\nu_N}}{d\bar{P}^{(N)}_{\nu_N}}(Q) = e^{Nh(Q)}.$$

Apply Sanov's theorem to the noninteracting system

Theorem

Let $\nu_N \to \nu$. Then the laws of the empirical measure for the noninteracting system $(\bar{P}^{(N)}_{\nu_N}, N \geq 1)$ satisfies the LDP in $\mathcal{M}_{1,\phi}(\mathcal{X})$ (with σ-field ... and topology ...) with speed N and rate function

$$J(Q) = \begin{cases} I(Q||\bar{P}) & \text{if } Q \circ \pi_0^{-1} = \nu \\ \infty & \text{otherwise,} \end{cases}$$

where $d\bar{P} = \sum_{z \in \mathcal{Z}} \nu(z) d\bar{P}_z$.

- Independent, but not identical because of possibly different initial conditions for particles.
- Use an extension provided by Dawson and Gartner.

Establish additional properties

- Whenever $J(Q) < \infty$, we have the following:
  - $Q \in \mathcal{M}_{1,\phi}(\mathcal{X})$, i.e., $\int \phi \, dQ < \infty$.
  - $Q \circ \pi_0^{-1} = \nu$
  - $h(\cdot)$ is continuous at Q
  - $\pi(\cdot)$ is continuous at Q
- Apply the Laplace-Varadhan principle: Since $(\bar{P}_{\nu_N}^{(N)}, N \geq 1)$, the law for empirical measure for the noninteracting system, satisfies an LDP, and since h is continuous at every point where $J(Q) < \infty$, argue that the interacting system's $(P_{\nu_N}^{(N)}, N \geq 1)$ satisfies the LDP with rate function

$$J(Q) - h(Q) = I(Q||P(\pi(Q))).$$

- h is not bounded. Its scaled cumulant is however bounded which suffices.
- Interpretation ...

Contraction principle

- We now have an LDP for empirical measures (Laws of Q_N).
  We want an LDP for the measure-valued process (Laws of $\mu_N(\cdot)$).

- Since $Q_N \overset{\pi}{\mapsto} \mu_N(\cdot)$ is continuous at all points where $J(Q) < \infty$, the push-forwards also satisfy the LDP with rate function:

$$S_{[0,T]}(\mu|\nu) = \inf\{I(Q||P(\pi(Q)) \mid \pi(Q) = \mu\}.$$

- Further calculations show that this is the same as the expression given before.

Final remarks for day 2

- In order to study large deviations from the fluid limit, we studied a larger object, empirical measure on path space.
- The steps:
  - Write the density of the interacting measure with respect to a noninteracting measure via Girsanov formula.
  - Apply Sanov's theorem for the noninteracting system.
  - Apply the Laplace-Varadhan principle for an LDP on the interacting system.
  - Contraction principle.
- Since we only assumed $\nu_N \to \nu$, but otherwise arbitrary initial conditions, we indeed have a stronger LDP that holds uniformly over the initial condition.
- The selection principle – coming soon via stationary distribution

Some exercises for students

- Consider a time-inhomogeneous jump Markov process $X(t)$ on the finite state space $\mathcal{Z}$ with transition rate matrix at time t given by $\Lambda(t)$. If the initial state $X(0)$ has distribution $\mu(0)$, how does $\mu(t)$ evolve over time? (Derive the forward equation).
- Let X be Bernoulli with parameter λ. Try to show that
$$\sup_{t\geq 0} \left[t \cdot \tau - \log \mathbb{E}[e^{tX}]\right] = I(\tau||\lambda).$$
- Suppose P be a Poisson point process on $[0, T]$ with intensity $\lambda(t)$. Let Q be the unit rate Poisson point process on $[0, T]$. Try to write the density of P with respect to Q at a realisation x, i.e., $\frac{dP}{dQ}(x)$. Take x to be a counting process with points at $t_1, t_2, \ldots, t_k$. (*Hint: Chop* $[0, T]$ *into disjoint intervals of duration h, use independent increments property, and let* $h \downarrow 0$).
- Try to prove the contraction principle: Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces, and $f : \mathcal{X} \to \mathcal{Y}$ a continuous function. Let $(X_N, N \geq 1)$ be a sequence of random variables on $\mathcal{X}$ that satisfy the LDP with speed N and good rate function I. Then $(f(X_N), N \geq 1)$ satisfies the LDP with speed N and good rate function $I'(y) = \inf\{I(x) \mid f(x) = y\}$.

References

1. Dawson, D. A. and Gärtner, J. (1987) "Large deviations from the McKean-Vlasov limit for weakly interacting diffusions", *Stochastics*, **20**, 247–308.
2. Feng, S. (1994) "Large deviations for Markov processes with mean field interaction and unbounded jumps", *Probability Th. and Related Fields* **100**, 227–252.
3. Feng, S. (1994) "Large deviations for empirical process of mean-field interacting particle system with unbounded jumps", *The Annals of Probability*, **22**(4), 2122–2151.
4. Djehiche, B. and Kaj, I. (1995) "The rate function for some measure-valued jump processes", *The Annals of Probability*, **23**(3), 1414–1438.
5. Léonard, C. (1995) "Large deviations for long range interacting particle systems with jumps", *Annales de l'I. H. P., section B*, **31**(2), 289–323.
6. Dai Pra, P. and den Hollander, F. (1996), "McKean-Vlasov limit for interacting random processes in random media", *Journal of Statistical Physics*, **84**(3/4), 735–772.
7. Freidlin, M. I. and Wentzell, A. D. (1998) *Random Perturbations of Dynamical Systems* (2nd ed.), Springer Verlag, Berlin–Heidelberg.
8. Del Moral, P. and Zajic, T. (2003) "Note on the Laplace-Varadhan Integral Lemma", *Bernoulli*, **9**(1), Feb. 2003, 49–65.

Recall from Lecture 1

- N-particle system with each particle's state coming from $\mathcal{Z}$, and with transitions in edge set $\mathcal{E}$.
- Transition rates are modulated by the mean-field. Rate matrix is $\Lambda(\mu_N(t))$.
- Kurtz's theorem: Let $\mu_N(0) \to \nu$ in probability. Then $\mu_N(\cdot) \to \mu(\cdot)$ in probability, uniformly over compacts. The fluid limit $\mu(\cdot)$ is the solution to the McKean-Vlasov equation

$$
\begin{aligned}
\dot{\mu}(t) &= \Lambda(\mu(t))^* \mu(t), \quad t \geq 0 \\
\mu(0) &= \nu.
\end{aligned}
$$

Standing assumptions

- The graph with vertex set $\mathcal{Z}$ and edge set $\mathcal{E}$ is irreducible
  Holds in the our WLAN case
- There exist positive constants $c > 0$ and $C < +\infty$ such that, for every $(i, j) \in \mathcal{E}$, we have

$$c \leq \lambda_{i,j}(\cdot) < C$$

- The mapping $\mu \mapsto \lambda_{i,j}(\mu)$ is Lipschitz continuous over $\mathcal{M}_1(\mathcal{Z})$

Recall from Lecture 3

- Theorem: Let $\mu_N(0) \to \nu$ (deterministic). Fix T. The sequence $(\mu_N(\cdot), N \geq 1)$ satisfies the LDP with speed N and good rate function

  $$S_{[0,T]}(\zeta(\cdot)|\nu) = \int_{[0,T]} \Big[\sum_{(i,j)\in\mathcal{E}} \zeta(t)(i)\lambda_{i,j}(\zeta(t))\tau^* \left(\frac{l_{i,j}(t)}{\lambda_{i,j}(\zeta(t))} - 1 \right) dt \Big]$$

  where $\dot{\zeta}(t) = L(t)^* \zeta(t)$.

- $S_{[0,T]}(\zeta(\cdot)|\nu)$: "resistance", cost of control $L(\cdot)$, cost of pushing the system along $\zeta(\cdot)$.

- The McKean-Vlasov path has cost 0.

Standing assumptions, and more

- The graph with vertex set $\mathcal{Z}$ and edge set $\mathcal{E}$ is irreducible
  Holds in the our WLAN case
- There exist positive constants $c > 0$ and $C < +\infty$ such that, for every $(i, j) \in \mathcal{E}$, we have

$$c \leq \lambda_{i,j}(\cdot) < C$$

- The mapping $\mu \mapsto \lambda_{i,j}(\mu)$ is Lipschitz continuous over $\mathcal{M}_1(\mathcal{Z})$
- (A) The McKean-Vlasov equation has ξ_0 as the globally asymptotically stable stationary point.
- Theorem: Under the above assumptions, $\mu_N(\infty) \to \xi_0$ in distribution (and hence in probability).

  (..., Stolyar 1989, Anantharam 1991, Anantharam and Benchekroun 1993, Bordenave et al. 2005/2007, Benaim and Le Boudec 2008)

The anticipated rate function

If $\mu_N(+\infty)$ is near ξ, then this is most likely due to an excursion that began at ξ_0, worked against the attractor ξ_0, and took the lowest cost path to ξ over all possible time durations.

LDP for the invariant measure (today)

Theorem

Under the same assumptions, $(\mu_N(\infty), N \geq 1)$ satisfies the LDP with speed N and rate function given as follows.

Looking backwards in time, consider the dynamics

$$\dot{\hat{\mu}}(t) = -\hat{L}(t)^* \hat{\mu}(t), t \geq 0$$

with $\hat{\mu}(0) = \xi$, $\lim_{t \to +\infty} \hat{\mu}(t) = \xi_0$, $\hat{L}(t)$ is some family of rate matrices, and $\hat{\mu}(t) \in \mathcal{M}_1(\mathcal{Z})$. The rate function is

$$s(\xi) = \inf_{\hat{\mu}} \int_{[0,+\infty)} \Big[\sum_{(i,j) \in \mathcal{E}} (\hat{\mu}(t)(i)) \lambda_{i,j}(\hat{\mu}(t)) \, \tau^* \left(\frac{\hat{l}_{i,j}(t)}{\lambda_{i,j}(\hat{\mu}(t))} - 1 \right) \Big] \, dt.$$

- We can also say, w.h.p., how the system arrived near ξ.

Generalisation: Freidlin-Wentzell theory

- Assumption (B): There exist a finite number of sets $K_1, K_2, \ldots, K_l$ (each compact) such that every ω-limit set of the McKean-Vlasov equation is a subset of one of the K_i.

Theorem

Under assumption (B), $(\mu_N(\infty), N \geq 1)$ satisfies the LDP with speed N and rate function

$$s(\xi) = \inf_i \inf_{\hat{\mu}} \left[s_i + \int_{[0,+\infty)} \Big[\sum_{(i,j)\in\mathcal{E}} (\hat{\mu}(t)(i))\lambda_{i,j}(\hat{\mu}(t))\ \tau^* \left(\frac{\hat{l}_{i,j}(t)}{\lambda_{i,j}(\hat{\mu}(t))} - 1 \right) \Big]\ dt \right]$$

where the second infimum is over all $\hat{\mu}$ that are solutions to $\dot{\hat{\mu}} = -\hat{L}(t)^*\hat{\mu}(t)$ for some family of rate matrices, initial condition $\hat{\mu}(0) = \xi$, terminal condition $\hat{\mu}(t) \to K_i$, and $\hat{\mu}(t) \in \mathcal{M}_1(\mathcal{Z})$ for all $t \geq 0$. The constants $s_1, \ldots, s_l$ are uniquely specified in terms of "resistances" to move between pairs of the compact sets.

Some general remarks

- The selection criterion. If there is a unique point at which s attains its minimum, then $\mu_N(\infty)$ tends to that point.
- Design system parameters to have a unique desired minimum point.

Proof steps (globally asymptotically stable equilibrium)

- Given $\nu_N \to \nu$, extract LDP for the laws for terminal state (finite T), via contraction principle, with rate function

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

- If the laws for initial states satisfy the LDP with a good rate function $s(\nu)$, argue that joint laws for initial and terminal states satisfy the LDP with a good rate function $s(\nu) + S_T(\xi|\nu)$. Then apply contraction principle to get that the laws for the terminal states satisfy the LDP with good rate function

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

- The invariant measures $(\wp^{(N)}, N \geq 1)$ live on a compact space. So, given any subsequence, there is a further subsequential LDP with appropriate speed, and with rate function $s(\xi)$ that satisfies

$$s(\xi) = \inf_{\nu \in \mathcal{M}_1(\mathcal{Z})} \{s(\nu) + S_T(\xi|\nu)\}$$

Proof steps continued

- By the assumption that ξ_0 is a unique equilibrium that is globally stable, we can show $s(\xi_0) = 0$.

- Extract a single infinite duration path $\hat{\mu}(\cdot)$ that is optimal, i.e., it attains the infimum for each duration $[0, mT]$, $\hat{\mu}(0) = \xi$, and satisfies

$$
\begin{aligned}
s(\xi) &= s(\hat{\mu}(mT)) + S_{mT}(\xi|\nu), \quad \forall m \geq 1 \\
&= s(\hat{\mu}(mT)) + \int_{[0,mT]} [\cdots] \, dt
\end{aligned}
$$

- The integrand in the second term is nonnegative; the second term increases with m, and so the first term $s(\hat{\mu}(mT))$ decreases with m. Since $s(\cdot)$ is bounded below by 0, $s(\hat{\mu}(mT))$ must converge to a constant as $m \to +\infty$

Proof steps continued even further

- So the increment $\int_{mT}^{mT+T}[\cdots]dt \to 0$ in the second term, and in the limit, integrand must be 0 a.e., which is a McKean-Vlasov path in reversed time.
  More precisely, $\hat{\mu}(\cdot)$ has an ω-limit set that is positively invariant to (McKean-Vlasov dynamics in reversed time)
  $$\hat{\mu}(t) = -\Lambda(\hat{\mu}(t))^* \hat{\mu}(t), \quad t \geq 0$$
- This limit set is also invariant to McKean-Vlasov dynamics. It is further compact and bounded within $\mathcal{M}_1(\mathcal{Z})$. The only such set invariant set is $\{\xi_0\}$. So $\hat{\mu}(mT) \to \xi_0$.
- Taking limit as $m \to +\infty$,
  $$s(\xi) = s(\xi_0) + \int_{[0,+\infty)} [\cdots] \; dt = 0 + \int_{[0,+\infty)} [\cdots] \; dt$$
  This expression is the same regardless of the initial subsequence
- Thus every subsequence has a further subsequence that satisfies the LDP with appropriate speed and the same rate function $s(\cdot)$.

Summary

- Mean-field model for a WLAN, and its fluid limit.
- A finite duration LDP for the measure-valued process.
- When there is a globally stable equilibrium ξ_0 for the McKean-Vlasov equation, the invariant measure satisfies the LDP. The rate function $s(\xi)$ is characterised by the cost of an optimal control that moves the system from ξ to ξ_0 in reversed time.
- Extension to cases with multiple stable points, and a selection criterion.

References

1. Borkar, V. S. and Sundaresan, R. (2012) "Asymptotics of the invariant measure in mean field models with jumps," *Stochastic Systems*, vol. 2, no. 2, pp.322–380. (arXiv: math.PR/1107.4142).
2. Freidlin, M. I. and Wentzell, A. D. (1998) *Random Perturbations of Dynamical Systems* (2nd ed.), Springer Verlag, Berlin–Heidelberg.