Mean-field Theory: Drift and the Mean Drift

Binan Gu

Department of Mathematical Sciences, New Jersey Institute of Technology

New Jersey Institute of Technology
Optimization & Machine Learning Seminar Spring 2022

New Jersey's Science & Technology University

Overview

Motivations

Wireless Sensor Networks

- Each sensor is a (random) vertex/node/agent.

Motivations

Wireless Sensor Networks

- Each sensor is a (random) vertex/node/agent.
- It collects and transmits data to the next node.

Motivations

Wireless Sensor Networks

- Each sensor is a (random) vertex/node/agent.
- It collects and transmits data to the next node.
- Large amount of nodes, so effective planning of transmission paths is important.

Motivations

Wireless Sensor Networks

- Each sensor is a (random) vertex/node/agent.
- It collects and transmits data to the next node.
- Large amount of nodes, so effective planning of transmission paths is important.

Example (Routing Algorithm)

Motivations

Wireless Sensor Networks

- Each sensor is a (random) vertex/node/agent.
- It collects and transmits data to the next node.
- Large amount of nodes, so effective planning of transmission paths is important.

Example (Routing Algorithm)

- Each node is associated with battery energy (and other constraints, e.g. storage, capacity, bandwidth, etc.).

Motivations

Wireless Sensor Networks

- Each sensor is a (random) vertex/node/agent.
- It collects and transmits data to the next node.
- Large amount of nodes, so effective planning of transmission paths is important.

Example (Routing Algorithm)

- Each node is associated with battery energy (and other constraints, e.g. storage, capacity, bandwidth, etc.).
- The goal is to maximize the lifetime of the network, i.e. preserving connectivity and preventing energy burnouts.

Motivations

Wireless Sensor Networks

- Each sensor is a (random) vertex/node/agent.
- It collects and transmits data to the next node.
- Large amount of nodes, so effective planning of transmission paths is important.

Example (Routing Algorithm)

- Each node is associated with battery energy (and other constraints, e.g. storage, capacity, bandwidth, etc.).
- The goal is to maximize the lifetime of the network, i.e. preserving connectivity and preventing energy burnouts.
- Data transmission between the nodes is like a queue, with random arrival and processing times.

Mean-field Theory

Mean-field Interaction Models

- Population processes in which individuals behave similarly.

Mean-field Theory

Mean-field Interaction Models

- Population processes in which individuals behave similarly.
- Care about “proportions of agents in a certain state" rather than the individual status of each agent.

Mean-field Theory

Mean-field Interaction Models

- Population processes in which individuals behave similarly.
- Care about “proportions of agents in a certain state" rather than the individual status of each agent.
- Symmetric reduction on the state space to gain efficiency in analysis.

Mean-field Interaction Models

- Population processes in which individuals behave similarly.
- Care about “proportions of agents in a certain state" rather than the individual status of each agent.
- Symmetric reduction on the state space to gain efficiency in analysis.
- Continuous and deterministic approximation of population processes in their **first moment**.

Mean-field Interaction Models

- Population processes in which individuals behave similarly.
- Care about “proportions of agents in a certain state" rather than the individual status of each agent.
- Symmetric reduction on the state space to gain efficiency in analysis.
- Continuous and deterministic approximation of population processes in their **first moment**.

Thomas G. Kurtz

Mean-field Theory

Mean-field Interaction Models

- Population processes in which individuals behave similarly.
- Care about “proportions of agents in a certain state" rather than the individual status of each agent.
- Symmetric reduction on the state space to gain efficiency in analysis.
- Continuous and deterministic approximation of population processes in their **first moment**.

Thomas G. Kurtz

- University of Wisconsin-Madison.

Mean-field Theory

Mean-field Interaction Models

- Population processes in which individuals behave similarly.
- Care about “proportions of agents in a certain state" rather than the individual status of each agent.
- Symmetric reduction on the state space to gain efficiency in analysis.
- Continuous and deterministic approximation of population processes in their **first moment**.

Thomas G. Kurtz

- University of Wisconsin-Madison.
- Convergence, approximation and representation of several important classes of Markov processes.

Mean-field Theory

Mean-field Interaction Models

- Population processes in which individuals behave similarly.
- Care about “proportions of agents in a certain state" rather than the individual status of each agent.
- Symmetric reduction on the state space to gain efficiency in analysis.
- Continuous and deterministic approximation of population processes in their **first moment**.

Thomas G. Kurtz

- University of Wisconsin-Madison.
- Convergence, approximation and representation of several important classes of Markov processes.
- Systems biology, population genetics, telecommunications networks and mathematical finance.

Example: Radio transmission in a single channel

Saturated nodes: every node always has a message to send. Each transmission has success probability depending on the number of nodes using the channel,

$$p_s(n) = 2^{-n}.$$

Two states:

1. Waiting to send, with p_1 to initiate.
2. Initiated but failed, and try to retransmit with p_2.

If transmission is successful in state 1, then the node stays in state 1 or otherwise goes to 2. Once retransmission is successful in state 2, it goes back to 1.

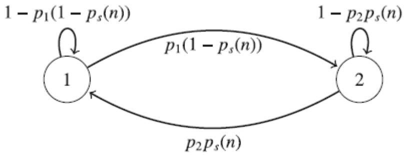

Fig. 1. The behaviour of a node in Example 1. The number of transmitting nodes is n.

Mean-Field Approximation

Agent Processes

Consider N agents with

- discrete time horizon $T = \mathbb{N}$.

Mean-Field Approximation

Agent Processes

Consider N agents with

- discrete time horizon $T = \mathbb{N}$.
- discrete state space $\mathcal{S}$.

Mean-Field Approximation

Agent Processes

Consider N agents with

- discrete time horizon $T = \mathbb{N}$.
- discrete state space $\mathcal{S}$.
- $\left\{X_i^{(N)}(t) : t \in T\right\}$, $\mathcal{S}$-valued discrete time Markov chains.

Mean-Field Approximation

Agent Processes

Consider N agents with

- discrete time horizon $T = \mathbb{N}$.
- discrete state space $\mathcal{S}$.
- $\left\{X_i^{(N)}(t) : t \in T\right\}$, $\mathcal{S}$-valued discrete time Markov chains.
- Each $X_i^{(N)}(t)$ follows a transition map $K_i : \mathcal{S}^N \times S \to [0, 1]$, namely, the probability $K_i(\mathbf{v}, s)$ that agent i transitions from
  - $\boldsymbol{v} \in \mathcal{S}^N$, state of the entire system (including agent i's current state) to
  - $s \in \mathcal{S}$, the next state.

Paradigms about Time Slots

Transition Types

Paradigms about Time Slots

Transition Types

1. Fully synchronized (i.e. synchronous DTMCs).

Paradigms about Time Slots

Transition Types

1. Fully synchronized (i.e. synchronous DTMCs).
2. Independent time slots (occurring at the same rate over sufficiently long intervals of time).

Paradigms about Time Slots

Transition Types

1. Fully synchronized (i.e. synchronous DTMCs).
2. Independent time slots (occurring at the same rate over sufficiently long intervals of time).

We discuss type 2, known as the *clock independence assumption*. This implies no simultaneous transitions (think of a property of the Poisson process). This assumption is not free.

Paradigms about Time Slots

Transition Types

1. Fully synchronized (i.e. synchronous DTMCs).
2. Independent time slots (occurring at the same rate over sufficiently long intervals of time).

We discuss type 2, known as the *clock independence assumption*. This implies no simultaneous transitions (think of a property of the Poisson process). This assumption is not free.

Let D be the time resolution (number of subintervals for one unit of time), and define $\epsilon = 1/D$ to form *global time*

$$T_G = \{0, \epsilon, 2\epsilon, \ldots\}.$$

Paradigms about Time Slots

Transition Types

1. Fully synchronized (i.e. synchronous DTMCs).
2. Independent time slots (occurring at the same rate over sufficiently long intervals of time).

We discuss type 2, known as the *clock independence assumption*. This implies no simultaneous transitions (think of a property of the Poisson process). This assumption is not free.

Let D be the time resolution (number of subintervals for one unit of time), and define $\epsilon = 1/D$ to form *global time*

$$T_G = \{0, \epsilon, 2\epsilon, \ldots\} .$$

The transition $K_i(\mathbf{v}, s)$ is then modified into:

$$\widehat{K}_i(\boldsymbol{v}, s) = \begin{cases} \epsilon K_i(\boldsymbol{v}, s), & \text{if } s \neq \boldsymbol{v}_i \\ 1 - \epsilon\left(1 - K_i(\boldsymbol{v}, s)\right), & \text{if } s = \boldsymbol{v}_i \end{cases}$$

Associated Stochastic Process

With the time-rescaled transition maps, we associate the collection of them with a new DTMC

$$Y^{(N)}(t) = \left(\widehat{X}_1^{(N)}(t), \ldots, \widehat{X}_N^{(N)}(t)\right)$$

with transition map (via independence of transitions)

$$\mathcal{K}^{(N)}(\boldsymbol{v}, \boldsymbol{v}') = \prod_{i=1}^{N} \widehat{K}_i(\boldsymbol{v}, \boldsymbol{v}'_i)$$

Definition (Mean Field Interaction Models (MFIM))

$Y^{(N)}(t)$ is an MFIM if

$$\mathcal{K}^{(N)}(\boldsymbol{v}, \boldsymbol{v}') = \mathcal{K}^{(N)}\left(\pi(\boldsymbol{v}), \pi(\boldsymbol{v}')\right)$$

for any permutation π on the agents.

Mean Field Interaction Models and Population Processes

Definition (Mean Field Interaction Models (MFIM))

$Y^{(N)}(t)$ is an MFIM if

$$\mathcal{K}^{(N)}(\boldsymbol{v}, \boldsymbol{v}') = \mathcal{K}^{(N)}\left(\pi(\boldsymbol{v}), \pi(\boldsymbol{v}')\right)$$

for any permutation π on the agents.

Note: this transition map may depend on the number agents in each state, but not on the state of a certain agent.

The Mean-Field Procedure

Occupation Measures

Proportion of Agents in State s

$$M_s^{(N)}(t) = \frac{1}{N} \sum_{1 \leq n \leq N} \mathbf{1}\left(\widehat{X}_n^{(N)}(t) = s\right)$$

whose values form an alternative representation Δ of the state space. The *agent model* $\left(\widehat{X}_1^{(N)}(t), M^{(N)}(t)\right) : t \in T_G$ is also Markov with (double) transition $P_1^{(N)}$.

Infinitesimal Generator for Arbitrary Agent

$$Q_{s,s'}^{(N)}(\boldsymbol{m}) = \lim_{D \to \infty} D P_{s,s'}^{(N)}(\boldsymbol{m})$$

where $P_{s,s'}^{(N)}(\boldsymbol{m})$ is the marginal transition probability of an arbitrary agent.

Drift and ODEs

Right-continuation with left limit (càdlàgs) of the process $M^{(N)}(t)$ via

$$\overline{M}^{(N)}(t) = M^{(N)}\left(\epsilon \lfloor Dt \rfloor\right).$$

Drift and ODEs

Right-continuation with left limit (càdlàgs) of the process $M^{(N)}(t)$ via

$$\overline{M}^{(N)}(t) = M^{(N)}\left(\epsilon \lfloor Dt \rfloor\right).$$

Know: $Q^{(N)}_{s,s'}(\boldsymbol{m})$, rate of any agent given current proportion $\boldsymbol{m}$.

Drift and ODEs

Right-continuation with left limit (càdlàgs) of the process $M^{(N)}(t)$ via

$$\overline{M}^{(N)}(t) = M^{(N)}\left(\epsilon \lfloor Dt \rfloor\right).$$

Know: $Q^{(N)}_{s,s'}(\boldsymbol{m})$, rate of any agent given current proportion $\boldsymbol{m}$.

Want: $\mathbb{E}\left[\overline{M}^{(N)}(t)\right]$, expected proportion of each state in time.

Drift and ODEs

Right-continuation with left limit (càdlàgs) of the process $M^{(N)}(t)$ via

$$\overline{M}^{(N)}(t) = M^{(N)}\left(\epsilon \lfloor Dt \rfloor\right).$$

Know: $Q^{(N)}_{s,s'}(\boldsymbol{m})$, rate of any agent given current proportion $\boldsymbol{m}$.

Want: $\mathbb{E}\left[\overline{M}^{(N)}(t)\right]$, expected proportion of each state in time.

Thus want: generator $F^{(N)}$ of $\overline{M}^{(N)}(t)$.

Drift and ODEs

Right-continuation with left limit (càdlàgs) of the process $M^{(N)}(t)$ via

$$\overline{M}^{(N)}(t) = M^{(N)}\left(\epsilon \lfloor Dt \rfloor\right).$$

Know: $Q^{(N)}_{s,s'}(\boldsymbol{m})$, rate of any agent given current proportion $\boldsymbol{m}$.

Want: $\mathbb{E}\left[\overline{M}^{(N)}(t)\right]$, expected proportion of each state in time.

Thus want: generator $F^{(N)}$ of $\overline{M}^{(N)}(t)$.
Procedure: prove a Dynkin-type formula

$$\mathbb{E}\left[\overline{M}^{(N)}(t) \mid \overline{M}^{(N)}(0)\right]$$
$$= \overline{M}^{(N)}(0) + \int_0^t \mathbb{E}\left[F^{(N)}\left(\overline{M}^{(N)}(s)\right) \mid \overline{M}^{N}(0)\right] ds$$

where a lot of ODE theory will apply (Picard, Gronwall, etc.)

Details

- Count number of transitions in $[t, t+\epsilon)$,

$$W_{s,s'}^{(N)}(t) = \sum_{k=1}^{N} \mathbf{1}\left\{\widehat{X}_k^{(N)}(t) = s,\ \widehat{X}_k^{(N)}(t+\epsilon) = s'\right\}$$

Details

- Count number of transitions in $[t, t+\epsilon)$,

$$W_{s,s'}^{(N)}(t) = \sum_{k=1}^{N} \mathbf{1}\left\{\widehat{X}_k^{(N)}(t) = s,\ \widehat{X}_k^{(N)}(t+\epsilon) = s'\right\}$$

- Express instantaneous change of the proportions,

$$\Delta\overline{M} = \overline{M}^{(N)}(t+\epsilon) - \overline{M}^{(N)}(t) = \sum_{s \neq s'} \frac{W_{s,s'}^{(N)}(t)}{N}(e_{s'} - e_s)$$

Details

- Count number of transitions in $[t, t+\epsilon)$,

$$W_{s,s'}^{(N)}(t) = \sum_{k=1}^{N} \mathbf{1}\left\{\widehat{X}_k^{(N)}(t) = s,\ \widehat{X}_k^{(N)}(t+\epsilon) = s'\right\}$$

- Express instantaneous change of the proportions,

$$\Delta\overline{M} = \overline{M}^{(N)}(t+\epsilon) - \overline{M}^{(N)}(t) = \sum_{s \neq s'} \frac{W_{s,s'}^{(N)}(t)}{N}(e_{s'} - e_s)$$

- Take an expectation, and then limit of rescale,

$$\widehat{F}^{(N)}(\boldsymbol{m}) = \mathbb{E}\left[\Delta\overline{M} \mid \overline{M}^{(N)}(t) = \boldsymbol{m}\right] = \sum_{s,s'} \boldsymbol{m}_s P_{s,s'}^{(N)}(\boldsymbol{m})(e_{s'} - e_s)$$

$$\boxed{F^{(N)}(\boldsymbol{m}) = \lim_{D \to \infty} D\widehat{F}^{(N)}(\boldsymbol{m}) = \sum_{s,s'} \boldsymbol{m}_s Q_{s,s'}^{(N)}(\boldsymbol{m})(e_{s'} - e_s)}$$

The Propagation of Chaos

Definition (ρ-chaotic Sequence)

Let $\rho \in \mathcal{M}(\mathcal{S})$ be a probability measure. For $N \geq 1$, the sequence $\{\rho_N\}$ of measures, each in $\mathcal{M}\left(\mathcal{S}^N\right)$, is ρ-chaotic iff for any fixed natural number k and bounded functions $\{f_i\}_{i=1}^{k}$:

$$\lim_{N \to \infty} \int_{\mathcal{S}^N} f_1(\boldsymbol{v}_1) f_2(\boldsymbol{v}_2) \ldots f_k(\boldsymbol{v}_k) \rho_N(d\boldsymbol{v}) = \prod_{i=1}^{k} \int_{\mathcal{S}} f_i(\boldsymbol{s}) \rho(d\boldsymbol{s})$$

i.e. a chaotic sequence mains a form of independence in the observations of separate agents in the limit. This is not utilized in the drift computation.

Mean Drift ODEs

One can show that $\overline{Y}^{(N)}(t)$, the joint agent process in continuous time, is μ-chaotic, where $\mu(i) = \mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]$ for $i \in \mathcal{S}$.

Mean Drift ODEs

One can show that $\overline{Y}^{(N)}(t)$, the joint agent process in continuous time, is μ-chaotic, where $\mu(i) = \mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]$ for $i \in \mathcal{S}$.

Asymptotic independence implies,

$$
\begin{aligned}
\mathbb{P}\left\{\overline{M}_i^{(N)}(t) = \frac{k}{N}\right\} &= \binom{N}{k}\text{Binomial stuff} \\
&\approx \binom{N}{k}\left(\mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]\right)^k\left(1 - \mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]\right)^{N-k} \\
&\approx e^{-N\mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]}\frac{\left(\mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]\right)^k}{k!}
\end{aligned}
$$

Mean Drift ODEs

One can show that $\overline{Y}^{(N)}(t)$, the joint agent process in continuous time, is μ-chaotic, where $\mu(i) = \mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]$ for $i \in \mathcal{S}$.

Asymptotic independence implies,

$$\mathbb{P}\left\{\overline{M}_i^{(N)}(t) = \frac{k}{N}\right\} = \binom{N}{k}\text{Binomial stuff}$$

$$\approx \binom{N}{k}\left(\mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]\right)^k \left(1 - \mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]\right)^{N-k}$$

$$\approx e^{-N\mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]} \frac{\left(\mathbb{E}\left[\overline{M}_i^{(N)}(t)\right]\right)^k}{k!}$$

Thus the generator can be computed using the asymptotically independent Poisson densities, extracting more information from the mere $Q_{s,s'}^{(N)}(\boldsymbol{m})$'s.

Example: Radio transmission in a single channel

Saturated nodes: every node always has a message to send. Each transmission has success probability depending on the number of nodes using the channel,

$$p_s(n) = 2^{-n}.$$

Two states:

1. Waiting to send, with p_1 to initiate.
2. Initiated but failed, and try to retransmit with p_2.

If transmission is successful in state 1, then the node stays in state 1 or otherwise goes to 2. Once retransmission is successful in state 2, it goes back to 1.

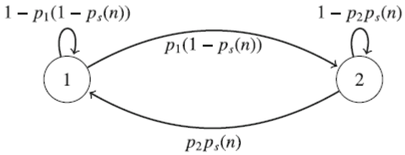

Fig. 1. The behaviour of a node in Example 1. The number of transmitting nodes is n.

Approximation comparisons

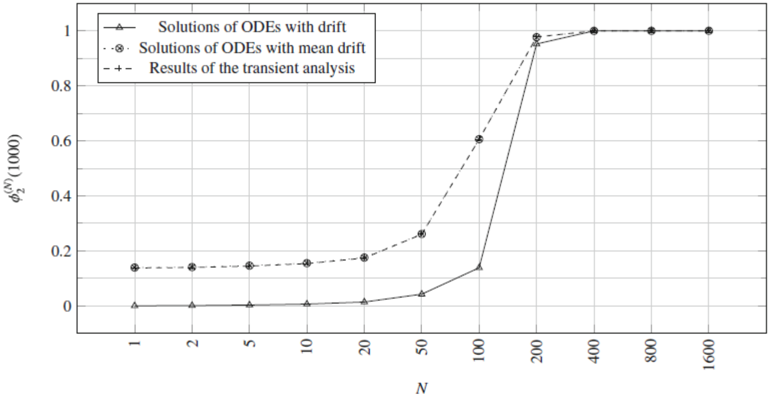

Fig. 2. Comparison between the proportion of nodes in the back-off state at time $t = 1000$ ($\phi_2^{(N)}(1000)$) for different network sizes N, based on a transient analysis of the Markov models (the solutions of the Chapman-Kolmogorov equations) and a mean field analysis by the ODEs incorporating the mean-drift.

This shows that the mean drift approximation is good even when *N* is not that large.

References

Mahmoud Talebi and Jan Friso Groote and Jean-Paul M.G. Linnartz “The Mean Drift: Tailoring the Mean Field Theory of Markov Processes for Real-World Applications ”. *Analytical and Stochastic Modelling Techniques and Applications*. 2017.