How to use mean field approximation for 10 players?

Nicolas Gast

Inria, Grenoble, France

ISCIS 2019 – Conférence Internationale sur les Méthodes de Modélisation des Systèmes et Réseaux Informatiques et Bio-Informatiques, October 2019

Discrete space mean field model

Population of N objects

- Each object evolves in a finite state-space $S_n(t) \in \mathcal{S}$.

Discrete space mean field model

Population of N objects

- Each object evolves in a finite state-space $S_n(t) \in \mathcal{S}$.

Evolution of *one object* : Markov kernel $Q(X)$.

$$X_i = \text{fraction of objects in state } i$$

$Q_{ij}(X) =$ rate/proba of one object of jumping from i to j.

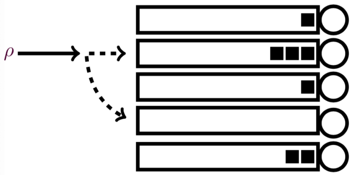
Q could represent:

- Game theory: Replicator dynamic, Best-response dynamics
- Biology: interactions between cells
- Computer Systems: decentralized allocations, cache management.

Some examples

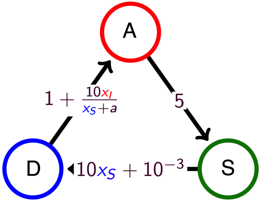
Load Balancing

(Mitzenmacher 98, Vvedenskaya 96)



Observe $d-1$ other nodes and chooses the shortest queue

Infection/information propagation
SIR / SIS



Mean field approximation

When the number of objects is large, objects become independent :

- In the synchronous case[1]:

$$X(t+1) = X(t)Q(X(t))$$

- In the asynchronous case[2].:

$$\frac{d}{dt}X(t) = X(t)Q(X(t))$$

In this talk, I will focus on the latter.

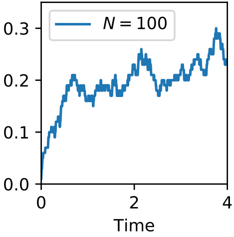
---

[1] Gomes, Mohr, Souza, 2010 : Discrete time, finite state space mean field games

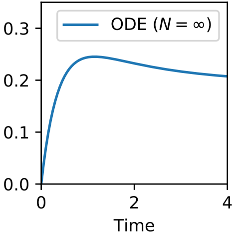
[2] Gomes,Mohr,Souza 2013: Continuous time finite state mean field game

This talk: compare finite N models and mean field approximation

$\lim_{N\to\infty}$



$=$



Mean field approximation $\dot{x} = xQ(x)$

$$\mathbf{P}\left[S_n(t) = i\right] \approx X_i(t) \approx x_i(t).$$

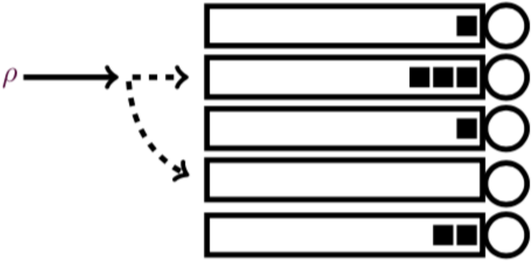
Outline

1. Classical Mean Field Limits
2. The Refined Mean Field
3. Extensions and Limits of the Approach

Outline

1. Classical Mean Field Limits
2. The Refined Mean Field
3. Extensions and Limits of the Approach

Example: the supermarket model ($SQ(d)$ load-balancing)



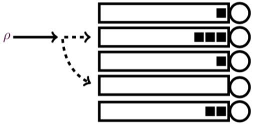
Arrival at each server ρ.

- Sample $d-1$ other queues.
- Allocate to the shortest queue

Service rate$=1$.

$SQ(d)$: state representation

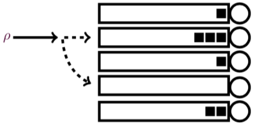
- Let $S_n(t)$ be the queue length of the nth queue at time t.



$$S = (1, 3, 1, 0, 2)$$

$SQ(d)$: state representation

- Let $S_n(t)$ be the queue length of the nth queue at time t.



$$S = (1, 3, 1, 0, 2)$$

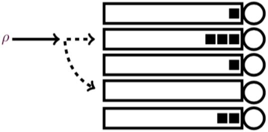
- Alternative representation:

$$X_i(t) = \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{\{S_n(t) \geq i\}},$$

which is the fraction of queues with queue length $\geq i$.

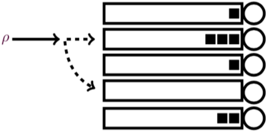
$$X = (1, 0.8, 0.4, 0.2, 0, 0, 0, \dots)$$

$SQ(d)$: state transitions



- Arrival: $x \mapsto x + \frac{1}{N}\mathbf{e_i}$.
- Departures: $x \mapsto x - \frac{1}{N}\mathbf{e_i}$.

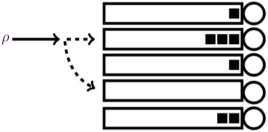
$SQ(d)$: state transitions



- Arrival: $x \mapsto x + \frac{1}{N}\mathbf{e_i}$.
- Departures: $x \mapsto x - \frac{1}{N}\mathbf{e_i}$.

Recall that x_i is the fraction of servers with i jobs or more. Pick two servers at random, what is the probability the least loaded has $i-1$ jobs?

$SQ(d)$: state transitions



- Arrival: $x \mapsto x + \frac{1}{N}\mathbf{e_i}$.
- Departures: $x \mapsto x - \frac{1}{N}\mathbf{e_i}$.

Recall that x_i is the fraction of servers with i jobs or more. Pick two servers at random, what is the probability the least loaded has $i-1$ jobs?

$x_{i-1}^2 - x_i^2$ when picked with replacement

$x_{i-1}\frac{Nx_{i-1}-1}{N-1} - x_i\frac{Nx_i - 1}{N-1}$ when picked without replacement

Note: this becomes asymptotically the same as N goes to infinity.

Transitions and mean field approximation

State changes on x:

$$x \mapsto x + \frac{1}{N}\mathbf{e_i} \text{ at rate } N\rho(x_{i-1}^d - x_i^d)$$
$$x \mapsto x - \frac{1}{N}\mathbf{e_i} \text{ at rate } N(x_i - x_{i+1})$$

The mean field approximation is to consider the ODE associated with the drift (average variation):

$$\dot{x}_i = \underbrace{\rho(x_{i-1}^d - x_i^d)}_{\text{Arrival}} - \underbrace{(x_i - x_{i+1})}_{\text{Departure}}$$

The model can be easily modified

Variants = push-pull model, centralized solution

- At rate r, each server that has $i \geq 2$ or more jobs probes a server and pushes a job to it if this server has 0 jobs. Transitions are:

$$x \mapsto x + \frac{1}{N}(-e_i + e_1) \text{ at rate } Nr(x_{i-1} - x_i)(1 - x_1)$$

The model can be easily modified

Variants = push-pull model, centralized solution

- At rate r, each server that has $i \geq 2$ or more jobs probes a server and pushes a job to it if this server has 0 jobs. Transitions are:

$$x \mapsto x + \frac{1}{N}(-e_i + e_1) \text{ at rate } Nr(x_{i-1} - x_i)(1 - x_1)$$

- At rate $N\gamma$, a centralized server serves a job from the longests queue. Transitions is:

$$x \mapsto x - \frac{1}{N}e_i \text{ at rate } N\gamma x_i \mathbf{1}_{\{x_{i+1}=0\}}$$

The model can be easily modified

Variants = push-pull model, centralized solution

- At rate r, each server that has $i \geq 2$ or more jobs probes a server and pushes a job to it if this server has 0 jobs. Transitions are:

$$x \mapsto x + \frac{1}{N}(-e_i + e_1) \text{ at rate } Nr(x_{i-1} - x_i)(1 - x_1)$$

- At rate $N\gamma$, a centralized server serves a job from the longests queue. Transitions is:

$$x \mapsto x - \frac{1}{N} e_i \text{ at rate } N\gamma x_i \mathbf{1}_{\{x_{i+1}=0\}}$$

The mean field approximation becomes (for $i > 1$):

$$\dot{x}_i = \underbrace{\rho(x_{i-1}^d - x_i^d)}_{\text{Arrival}} - \underbrace{(x_i - x_{i+1})}_{\text{Departure}} - \underbrace{r(x_{i-1} - x_i)(1 - x_1)}_{\text{Push}} - \underbrace{N\gamma x_i \mathbf{1}_{\{x_{i+1}=0\}}}_{\text{Centralized}}$$

$$\dot{x}_1 = \underbrace{\rho(x_0^d - x_1^d)}_{\text{Arrival}} - \underbrace{(x_1 - x_2)}_{\text{Departure}} + \sum_{i=2}^{\infty} \underbrace{r(x_{i-1} - x_i)(1 - x_1)}_{\text{Push}} - \underbrace{N\gamma x_1 \mathbf{1}_{\{x_2=0\}}}_{\text{Centralized}}$$

These models are examples of density dependent population processes (Introduced by (Kurtz, 70s))

A population process is a sequence of CTMCs $X^N(t)$ indexed by the population size N, with state space $E^N \subset E$ and transitions (for $\ell \in \mathcal{L}$):

$$X \mapsto X + \frac{\ell}{N} \qquad \text{at rate } N\beta_\ell(X).$$

These models are examples of density dependent population processes (Introduced by (Kurtz, 70s))

A population process is a sequence of CTMCs $X^N(t)$ indexed by the population size N, with state space $E^N \subset E$ and transitions (for $\ell \in \mathcal{L}$):

$$X \mapsto X + \frac{\ell}{N} \qquad \text{at rate } N\beta_\ell(X).$$

The drift is $f(x) = \sum \ell\beta_\ell(x)$ and the mean field approximation is the solution of the ODE

$$\dot{x} = f(x).$$

These models are examples of density dependent population processes (Introduced by (Kurtz, 70s))

A population process is a sequence of CTMCs $X^N(t)$ indexed by the population size N, with state space $E^N \subset E$ and transitions (for $\ell \in \mathcal{L}$):

$$X \mapsto X + \frac{\ell}{N} \qquad \text{at rate } N\beta_\ell(X).$$

The drift is $f(x) = \sum \ell\beta_\ell(x)$ and the mean field approximation is the solution of the ODE

$$\dot{x} = f(x).$$

Example: SQ(d) load balancing: $\dot{x}_i = \rho(x_{i-1}^d - x_i^d) - (x_i - x_{i+1})$. This ODE has a unique attractor: $\pi_i = \rho^{(d^i-1)/(d-1)}$.

Convergence result as N goes to infinity

Theorem (under some mild conditions, mostly Lipschitz continuity): If $X^N(0)$ converges to x_0, then for any finite T:

$$\sup_{0 \le t \le T} \left\| X^N(t) - x(t) \right\| \to 0.$$

where $x(t)$ is the unique solution of the ODE $\dot{x} = f(x)$.

Convergence result as N goes to infinity

Theorem (under some mild conditions, mostly Lipschitz continuity): If $X^N(0)$ converges to x_0, then for any finite T:

$$\sup_{0\leq t\leq T} \left\| X^N(t) - x(t) \right\| \to 0.$$

where $x(t)$ is the unique solution of the ODE $\dot{x} = f(x)$.

Theorem If the mean field approximation as a unique attractor $x(\infty)$, then

$$\left\| X^N(\infty) - x(\infty) \right\| \to 0$$

SQ(d) load balancing ($d = 2$)

N	Simulation (steady-state ave. queue length) 10	20	50	100	Fixed point ∞ (mean field)
$\rho = 0.70$	1.2194	1.1735	1.1471	1.1384	1.1301
$\rho = 0.90$	2.8040	2.5665	2.4344	2.3931	2.3527
$\rho = 0.95$	4.2952	3.7160	3.4002	3.3047	3.2139

Fairly good accuracy for $N = 100$ servers.

SQ(d) load balancing ($d = 2$)

	Simulation (steady-state ave. queue length)				Fixed point
N	10	20	50	100	∞ (mean field)
$\rho = 0.70$	1.2194	1.1735	1.1471	1.1384	1.1301
$\rho = 0.90$	2.8040	2.5665	2.4344	2.3931	2.3527
$\rho = 0.95$	4.2952	3.7160	3.4002	3.3047	3.2139

Fairly good accuracy for $N = 100$ servers.

Pull-push model (servers with ≥ 2 jobs push to empty)

N	Simulation (steady-state ave. queue length) 10	20	50	100	Fixed point ∞ (mean field)
$\rho = 0.80$	1.5569	1.4438	1.3761	1.3545	1.3333
$\rho = 0.90$	2.3043	1.9700	1.7681	1.7023	1.6364
$\rho = 0.95$	3.4288	2.6151	2.1330	1.9720	1.8095

Fairly good accuracy for $N = 100$ servers.

Outline

1 Classical Mean Field Limits

2 The Refined Mean Field

3 Extensions and Limits of the Approach

Mean Field Accuracy

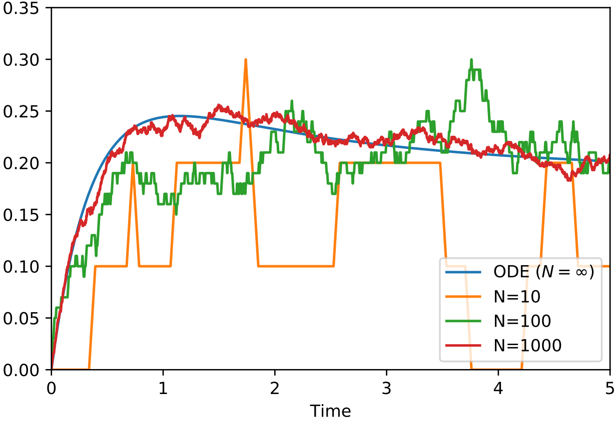
Theorem (Kurtz (1970s), Ying (2016)):

If the drift f is Lipschitz-continuous:

$$X^N(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$

If in addition the ODE has a unique attractor π:

$$\mathbb{E}\left[X^N(\infty) - \pi\right] = O(1/\sqrt{N})$$



Expected values estimated by mean field are 1/N-accurate

Some experiments (for SQ(2) with $\rho = 0.9$):

N	10	100	1000	∞
Average queue length (simulation)	2.8040	2.3931	2.3567	2.3527
Error of mean field	0.4513	0.0404	0.0040	0

Expected values estimated by mean field are 1/N-accurate

Some experiments (for SQ(2) with $\rho = 0.9$):

N	10	100	1000	∞
Average queue length (simulation)	2.8040	2.3931	2.3567	2.3527
Error of mean field	0.4513	0.0404	0.0040	0

Error seems to decrease as $1/N$

Expected values estimated by mean field are 1/N-accurate

Some experiments (for SQ(2) with $\rho = 0.9$):

N	10	100	1000	∞
Average queue length (simulation)	2.8040	2.3931	2.3567	2.3527
Error of mean field	0.4513	0.0404	0.0040	0

Error seems to decrease as $1/N$

Theorem (Kolokoltsov 2012, G. 2017& 2018). If the drift f is C^2 and has a unique exponentially stable attractor, then for any $t \in [0, \infty) \cup \{\infty\}$, there exists a constant V_t such that:

$$\mathbb{E}\left[h(X^N(t))\right] = h(x(t)) + \frac{V(t)}{N} + O(1/N^2)$$

The refined mean field approximation...

... is defined as the classic mean field plus the $1/N$ correction term:

$$\mathbb{E}\left[X^N\right] = \underbrace{x(t) + \frac{V(t)}{N}}_{\text{Refined mf approx}} + O(1/N^2),$$

where $V(t)$ is computed analytically.

The refined mean field approximation...

... is defined as the classic mean field plus the $1/N$ correction term:

$$\mathbb{E}\left[X^N\right] = \underbrace{x(t) + \frac{V(t)}{N}}_{\text{Refined mf approx}} + O(1/N^2),$$

where $V(t)$ is computed analytically.

To compute $V(t)$, we need:

- Derivative of the drifts:

$$F^i_j(t) = \frac{\partial f_i}{\partial x_j}(x(t)) \text{ and } F^i_{jk}(t) = \frac{\partial^2 f_i}{\partial x_j \partial x_k}(x(t))$$

- A variance term:

$$Q(t) = \sum_{\ell} \ell \otimes \ell \beta_{\boldsymbol{\ell}}(X(t))$$

Computational methods

Theorem (G, Van Houdt 2018) Given a density dependent process with twice-differentiable drift. Let $h : E \to \mathbb{R}$ be a twice-differentiable function, then for $t > 0$:

$$\mathbb{E}\left[h(X^N(t))\right] = h(x(t)) + \frac{1}{N}\left(\sum_i \frac{\partial h(x(t))}{\partial x_i} V_i(t) + \frac{1}{2}\sum_{ij} \frac{h(x(t))}{\partial x_i \partial x_j} W_{ij}(t)\right) + O(\frac{1}{N^2}),$$

where

$$\frac{d}{dt} V^i = \sum_j F^i_j V^j + \sum_{jk} F^i_{j,k} W^{j,k}$$

$$\frac{d}{dt} W^{j,k} = Q^{jk} + \sum_m F^j_m W^{m,k} + \sum_m W^{j,m} F^k_m$$

Theorem (G, Van Houdt 2018) The previous theorem also holds for the stationary regime ($t = +\infty$) if the ODE has a unique exponentially stable attractor.

The supermarket model (SQ(2))

N	10	20	30	50	100	∞
$\rho = 0.7$						
Simulation	1.2194	1.1735	1.1584	1.1471	1.1384	–
Refined mf	1.2150	1.1726	1.1584	1.1471	1.1386	1.1301
$\rho = 0.9$						
Simulation	2.8040	2.5665	2.4907	2.4344	2.3931	–
Refined mf	2.7513	2.5520	2.4855	2.4324	2.3925	2.3527
$\rho = 0.95$						
Simulation	4.2952	3.7160	3.5348	3.4002	3.3047	–
Refined mf	4.1017	3.6578	3.5098	3.3915	3.3027	3.2139

Average queue length: Refined mean field approximation gives a significant improvement.

The supermarket model (SQ(2))

N	10	20	30	50	100	∞
$\rho = 0.7$						
Simulation	1.2194	1.1735	1.1584	1.1471	1.1384	–
Refined mf	1.2150	1.1726	1.1584	1.1471	1.1386	1.1301
$\rho = 0.9$						
Simulation	2.8040	2.5665	2.4907	2.4344	2.3931	–
Refined mf	2.7513	2.5520	2.4855	2.4324	2.3925	2.3527
$\rho = 0.95$						
Simulation	4.2952	3.7160	3.5348	3.4002	3.3047	–
Refined mf	4.1017	3.6578	3.5098	3.3915	3.3027	3.2139

Average queue length: Refined mean field approximation gives a significant improvement.

The supermarket model (SQ(2))

N	10	20	30	50	100	∞
$\rho = 0.7$						
Simulation	1.2194	1.1735	1.1584	1.1471	1.1384	–
Refined mf	1.2150	1.1726	1.1584	1.1471	1.1386	1.1301
$\rho = 0.9$						
Simulation	2.8040	2.5665	2.4907	2.4344	2.3931	–
Refined mf	2.7513	2.5520	2.4855	2.4324	2.3925	2.3527
$\rho = 0.95$						
Simulation	4.2952	3.7160	3.5348	3.4002	3.3047	–
Refined mf	4.1017	3.6578	3.5098	3.3915	3.3027	3.2139

Mean field approximation

Average queue length: Refined mean field approximation gives a significant improvement.

Pull-push model (servers with ≥ 2 jobs push to empty)

N	10	20	50	100	∞
$\rho = 0.8$					
Simulation	1.5569	1.4438	1.3761	1.3545	–
Refined mean field	1.5473	1.4403	1.3761	1.3547	1.3333
$\rho = 0.90$					**Mean field approximation**
Simulation	2.3043	1.9700	1.7681	1.7023	–
Refined mean field	2.2945	1.9654	1.7680	1.7022	1.6364
$\rho = 0.95$					
Simulation	3.4288	2.6151	2.1330	1.9720	–
Refined mean field	3.4369	2.6232	2.1350	1.9723	1.8095

Average queue length: Refined mean field approximation is remarkably accurate

Outline

1 Classical Mean Field Limits

2 The Refined Mean Field

3 Extensions and Limits of the Approach

Recap and extensions

If $x \mapsto xQ(x)$ is C^2, then :

1. The accuracy of the classical mean field approximation is $O(1/N)$.
2. We can use this to define a refined approximation.
3. The refined approximation is often accurate for $N = 10$:

Recap and extensions

If $x \mapsto xQ(x)$ is C^2, then :

1. The accuracy of the classical mean field approximation is $O(1/N)$.
2. We can use this to define a refined approximation.
3. The refined approximation is often accurate for $N = 10$:

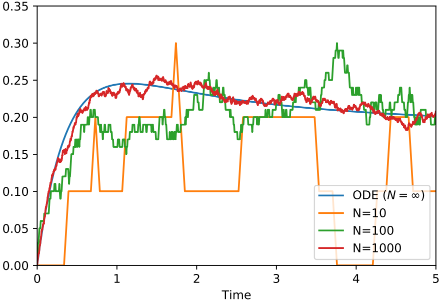
Extensions:

- Transient regime
- Discrete-time systems
- We can also compute the next term in $1/N^2$.

Limit 1: it applies to object properties but not to populations

Population's state: $X(t) = \frac{1}{N}\sum_{n=1}^{N} \delta_{S_n(t)}$

$$X(t) = x(t) + \frac{G(t)}{\sqrt{N}}$$



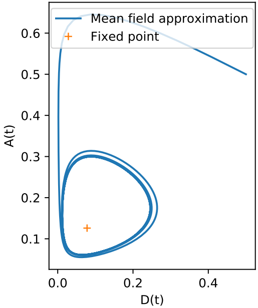
One object has state $S_n(t)$

$$\mathbb{E}[X(t)] = x(t) + \frac{C}{N}$$

Average queue length ($N = 10$ and $\rho = 0.9$)

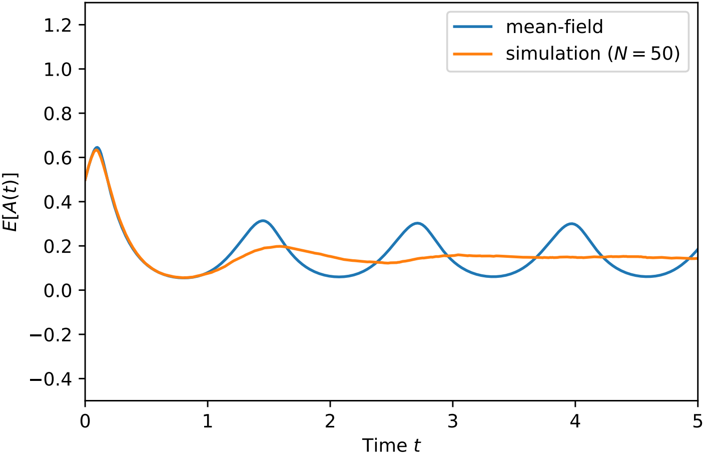
Simu	Refined M.F.	M.F.
2.804	2.751	2.353

Limit 2: It can fail when the mean field approximation has limiting cycles

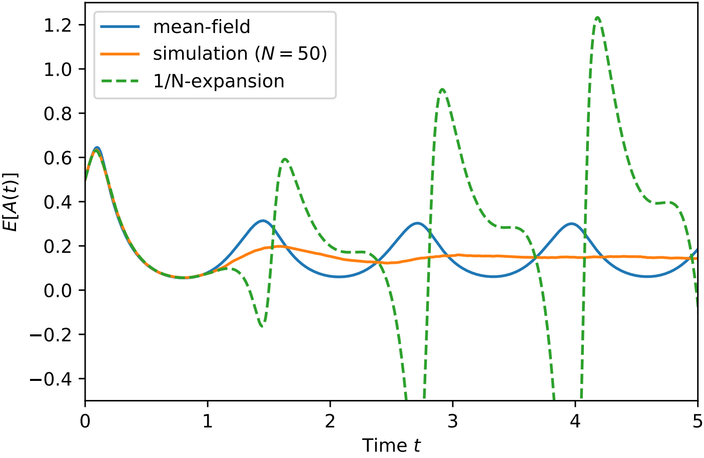


Transition	Rate
$(D, A, S) \mapsto (D - \frac{1}{N}, A + \frac{1}{N}, S)$	$N(0.1 + 10X_A)X_D$
$(D, A, S) \mapsto (D, A - \frac{1}{N}, S + \frac{1}{N})$	$N5X_A$
$(D, A, S) \mapsto (D + \frac{1}{N}, A, S - \frac{1}{N})$	$N(1 + \frac{10X_A}{X_D + \delta})X_S$

Limit 2: It can fail when the mean field approximation has limiting cycles



Limit 2: It can fail when the mean field approximation has limiting cycles



Limit 3: What about games and/or optimal control?

Discrete-state mean field games are relatively "easy" to work with.

- Forward equation : ODE.
- Backward equation : MDP (Markov decision process)

Open question : Do the Nash equilibria of the finite games converge to a mean field equilibria? What is the rate of convergence?

Limit 3: What about games and/or optimal control?

Discrete-state mean field games are relatively “easy” to work with.

- Forward equation : ODE.
- Backward equation : MDP (Markov decision process)

Open question : Do the Nash equilibria of the finite games converge to a mean field equilibria? What is the rate of convergence?

- The value of the game does not always converge (Doncel et al. 2017)
- When it does, convergence seems to be $O(1/\sqrt{N})$.

Some References

http://polaris.imag.fr/nicolas.gast

nicolas.gast@inria.fr

- A Refined Mean Field Approximation by Gast and Van Houdt. SIGMETRICS 2018 (best paper award)
- Size Expansions of Mean Field Approximation: Transient and Steady-State Analysis Gast, Bortolussi, Tribastone
- Expected Values Estimated via Mean Field Approximation are O(1/N)-accurate by Gast. SIGMETRICS 2017.
- https://github.com/ngast/rmf_tool/