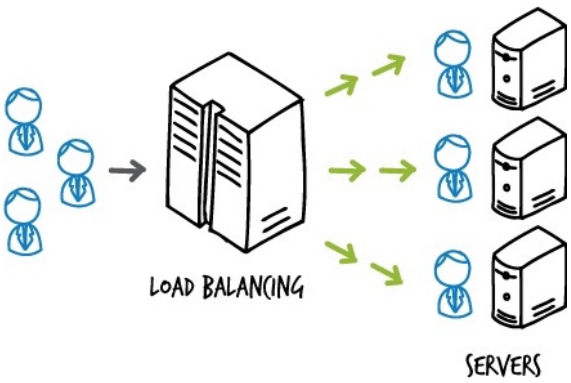
The bias of mean field approximation

Nicolas Gast (Inria, Grenoble)

joint work with Sebastian Allmeier (Inria)

Séminaire Univ. Laval (Québec) April 2022

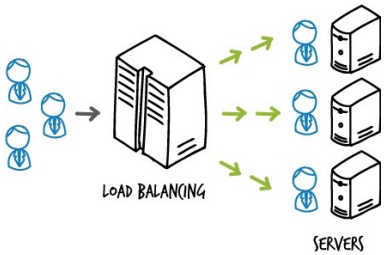
Motivation: Studying interacting particle systems



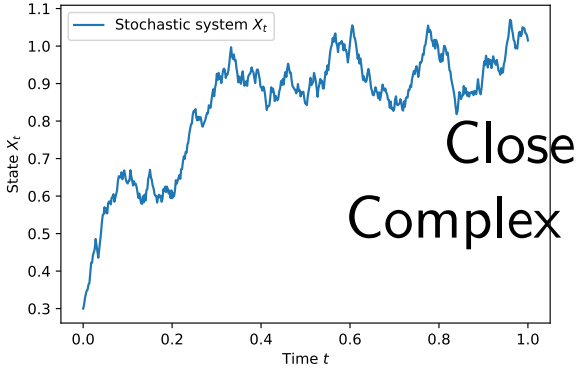
- Stochastic models are complex.

Fluid / mean field approximation simplifies the analysis

n servers



stochastic model

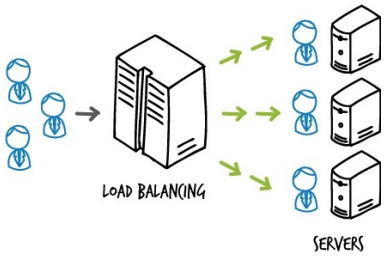


Close form?

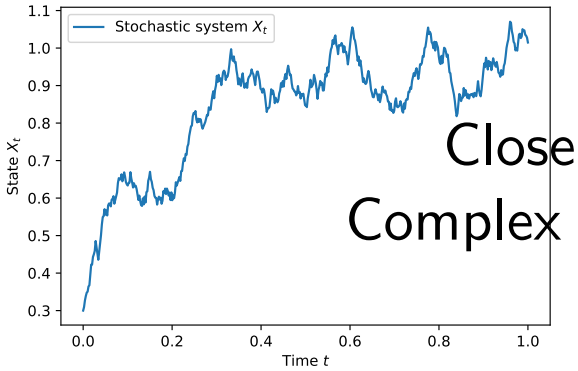
Complex to analyze

Fluid / mean field approximation simplifies the analysis

n servers

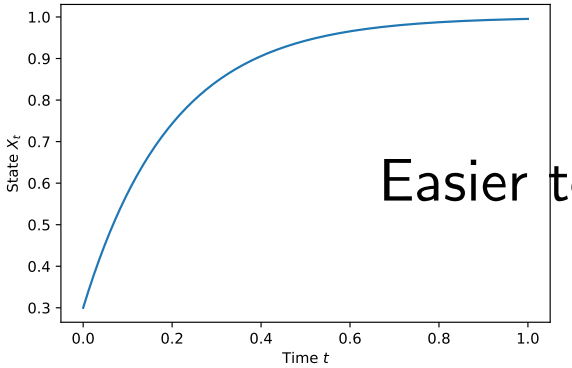


stochastic model



Close form?
Complex to analyze

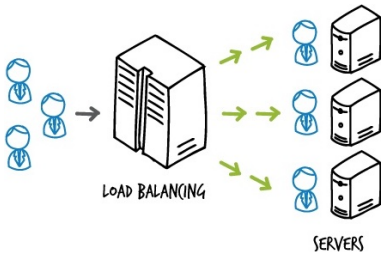
fluid



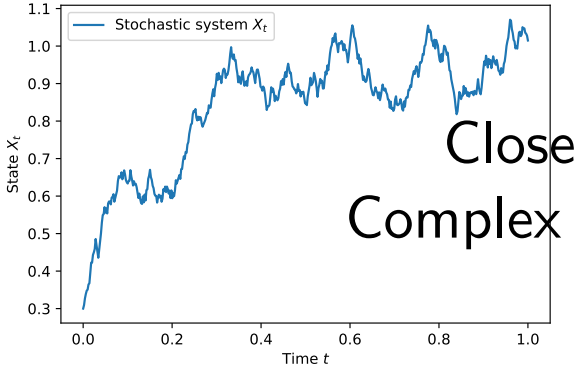
Easier to analyze

Fluid / mean field approximation simplifies the analysis

n servers



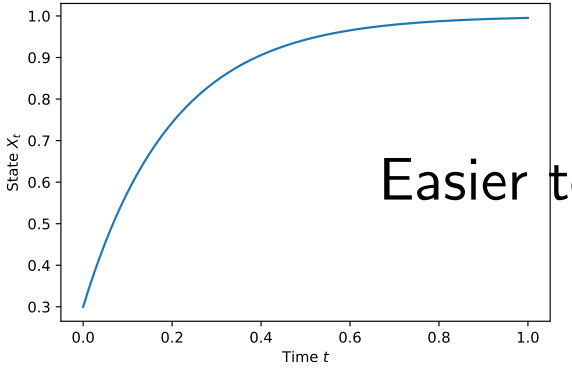
stochastic model



Close form?
Complex to analyze

Question: What is the error made?

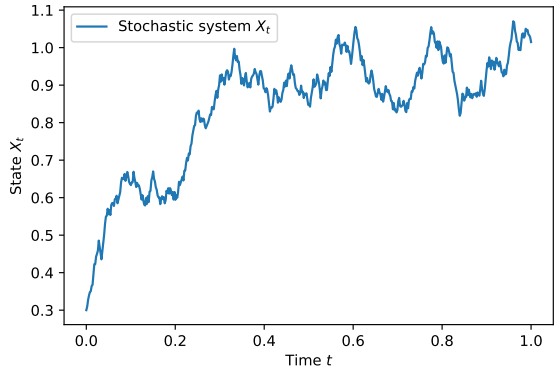
fluid



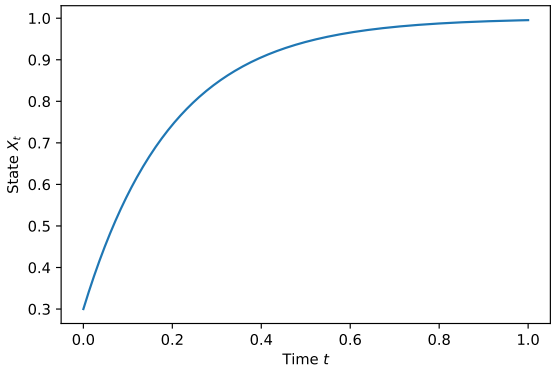
Easier to analyze

Fluid approximation is often justified by a law of large numbers

$$\lim_{n\to\infty}$$



$$=$$



almost surely.

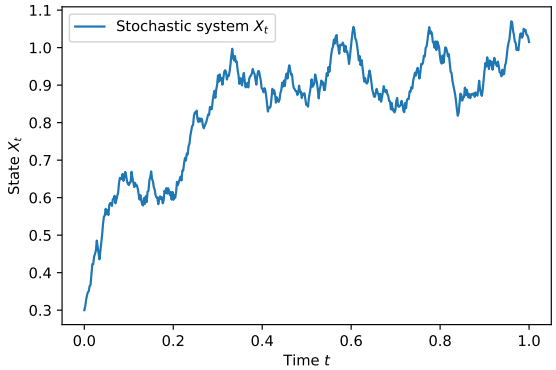
- Bound between X_t and $\phi_t(X_0)$ by using Gronwall's lemma.

$$X_t - X_0 - \int_0^t f(X_s)ds \text{ is a martingale.}$$

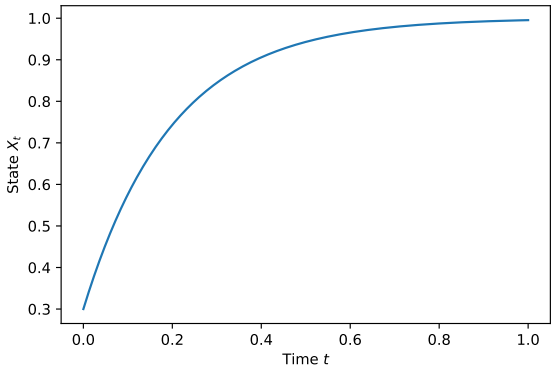
- This gives a $O(1/\sqrt{n})$ convergence-rate.

Fluid approximation is often justified by a law of large numbers

$$\lim_{n\to\infty}$$



$$=$$



almost surely.

- Bound between X_t and $\phi_t(X_0)$ by using Gronwall's lemma.

$$X_t - X_0 - \int_0^t f(X_s)ds \text{ is a martingale.}$$

- This gives a $O(1/\sqrt{n})$ convergence-rate.

Talk:

- Explain mean-field approximation through examples.
- Show tools to provide sharp convergence results.

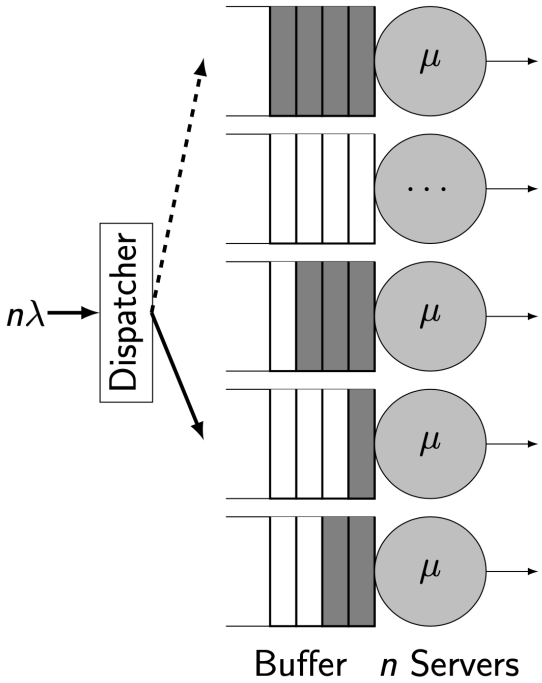
(Main) Related work

- Kurtz, 70s.
  - Fluid limits, diffusion limits (mostly transient regime)
- Application to queues
  - Fluid limits (Bramson, Dai 90s)
  - Interacting queues and mean-field: Load balancing (Mitzenmacher 01 + many recent)
- Stein's method:
  - Stein (1986)
  - Application to queueing: Braverman, Dai (2017–)
  - Application to mean-field models: Ying (2017).
- Refined mean field / Size expansions
  - Computational biology: Grima et al (2010s)
  - G. Van Houdt (2018), Allmeier G. (2021,2022).

Outline

Example: SQ(2) model

Dispatcher sends to the shortest among two random queues.



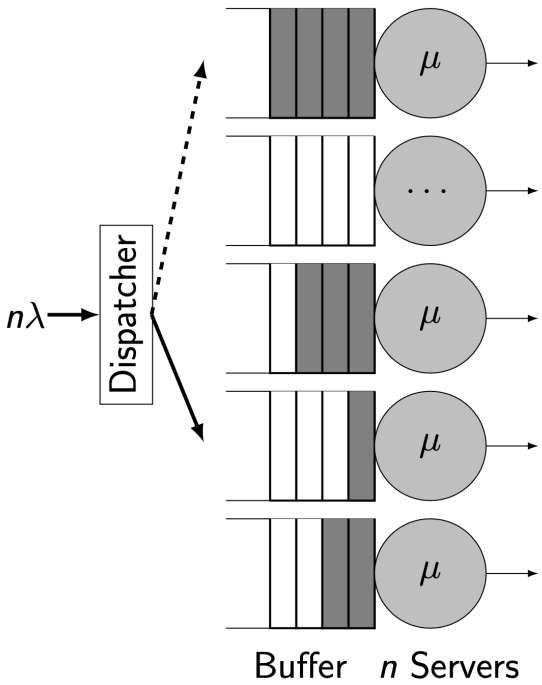
$\mathbf{Q} = (4, 0, 3, 1, 2)$.

Natural Markov model: $(Q_1 \ldots Q_n)$.

- Complexity grows with n.

Example: SQ(2) model

Dispatcher sends to the shortest among two random queues.



$\mathbf{Q} = (4, 0, 3, 1, 2)$.

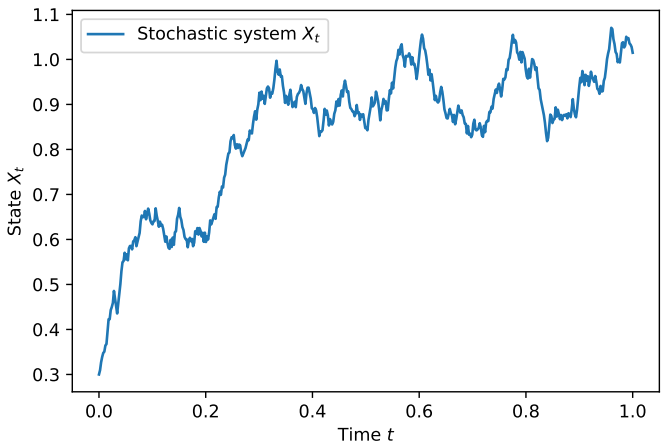
$\mathbf{X} = (1, 0.8, 0.6, 0.4, 0.2, 0, \ldots)$.

Natural Markov model: $(Q_1 \ldots Q_n)$.

- Complexity grows with n.

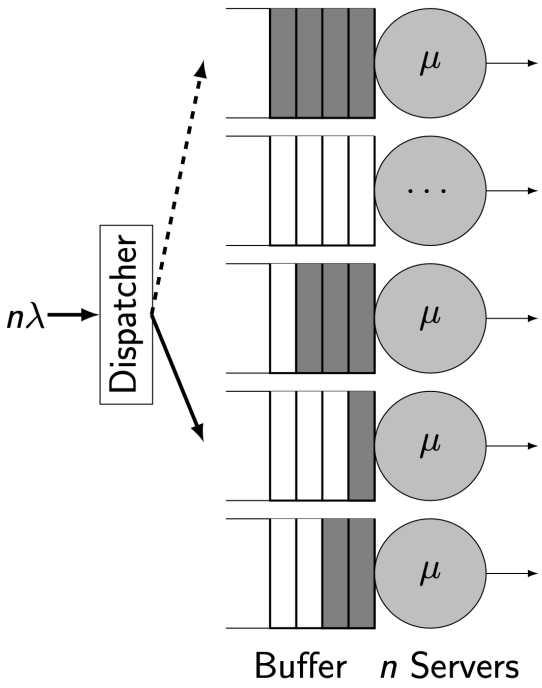
"Simplified" process: "empirical measure"

$$X_i = \frac{1}{n}\#\{\text{queue with } i \text{ or more}\}.$$



Example: SQ(2) model

Dispatcher sends to the shortest among two random queues.



$\mathbf{Q} = (4, 0, 3, 1, 2)$.

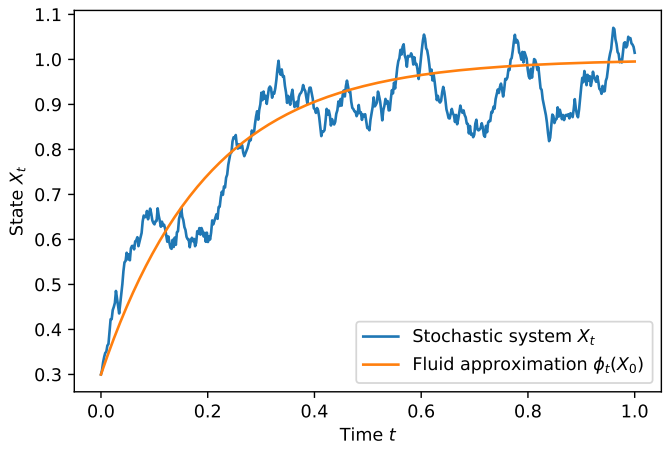
$\mathbf{X} = (1, 0.8, 0.6, 0.4, 0.2, 0, \ldots)$.

Natural Markov model: $(Q_1 \ldots Q_n)$.

- Complexity grows with n.

"Simplified" process: "empirical measure"

$$X_i = \frac{1}{n}\#\{\text{queue with } i \text{ or more}\}.$$



How to construct the fluid approximation?

(=hydrodynamic limit)

The transitions on X_i are:

$$X_i \mapsto X_i - \frac{1}{n} \qquad \text{at rate } n\mu(X_i - X_{i+1}) \qquad \textit{// job completion}$$

$$X_i \mapsto X_i + \frac{1}{n} \qquad \text{at rate } n\lambda(X_{i-1}^2 - X_i^2) \qquad \textit{// job arrival}$$

The ODE is $\dot{x}_i = \mu(x_{i+1} - x_i) + \lambda(x_{i-1}^2 - x_i^2)$.

How to construct the fluid approximation?

(=hydrodynamic limit)

The transitions on X_i are:

$$X_i \mapsto X_i - \frac{1}{n} \qquad \text{at rate } n\mu(X_i - X_{i+1}) \qquad \textit{// job completion}$$

$$X_i \mapsto X_i + \frac{1}{n} \qquad \text{at rate } n\lambda(X_{i-1}^2 - X_i^2) \qquad \textit{// job arrival}$$

The ODE is $\dot{x}_i = \mu(x_{i+1} - x_i) + \lambda(x_{i-1}^2 - x_i^2)$.

The fixed point is a good approximation of the average queue length:

n	10	100	1000	Fixed point
Average queue length, $\lambda/\mu = 0.9$	2.804	2.393	2.357	2.353

How to construct the fluid approximation?

(=hydrodynamic limit)

The transitions on X_i are:

$$X_i \mapsto X_i - \frac{1}{n} \qquad \text{at rate } n\mu(X_i - X_{i+1}) \qquad \textit{// job completion}$$

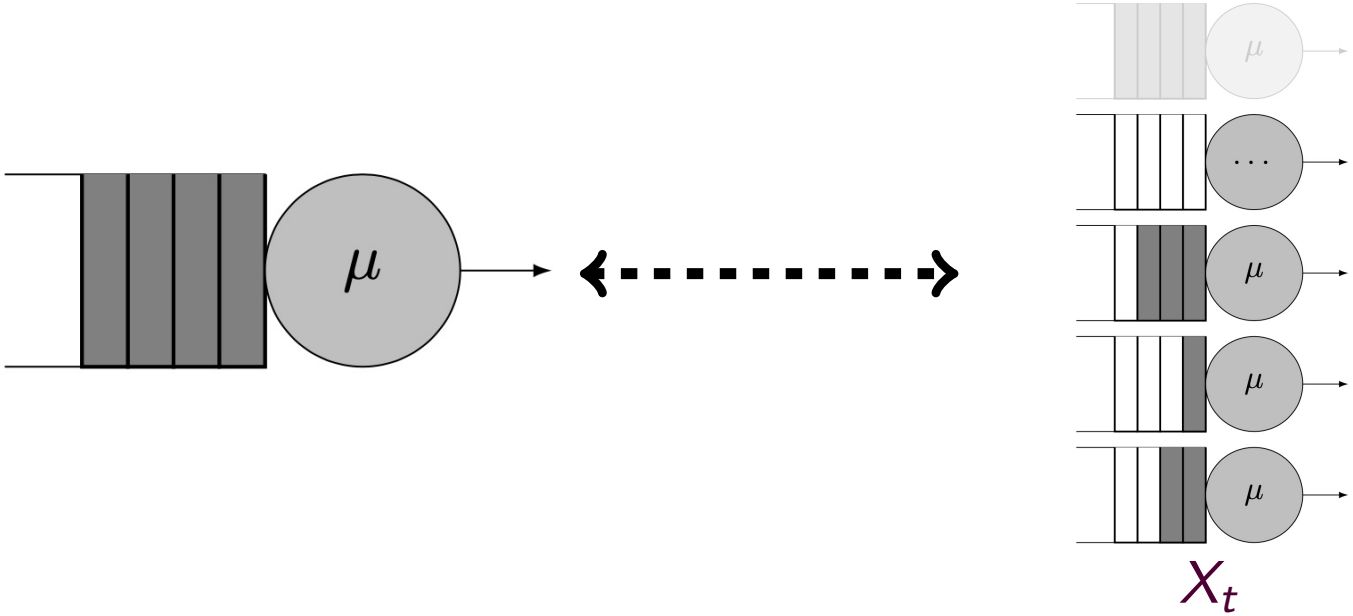
$$X_i \mapsto X_i + \frac{1}{n} \qquad \text{at rate } n\lambda(X_{i-1}^2 - X_i^2) \qquad \textit{// job arrival}$$

The ODE is $\dot{x}_i = \mu(x_{i+1} - x_i) + \lambda(x_{i-1}^2 - x_i^2)$.

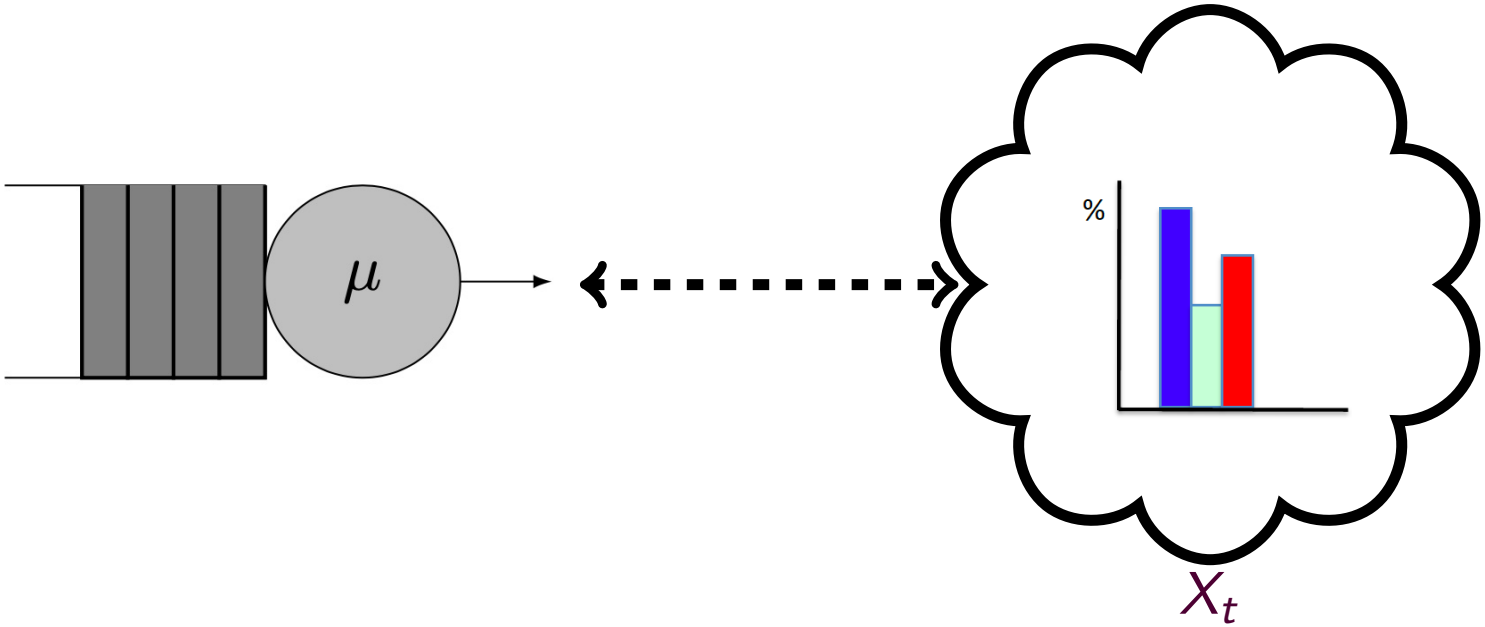
The fixed point is a good approximation of the average queue length:

n	10	100	1000	Fixed point
Average queue length, $\lambda/\mu = 0.9$	2.804	2.393	2.357	2.353
Bias	0.45	0.039	0.004	0

Why is this approximation called a "mean field approximation" (as in physics)



Why is this approximation called a "mean field approximation" (as in physics)



- X_t is the law of the queue.
- The queue interacts with its law (McKean-Vlasov process).

This corresponds to assuming that queues are independent.

This extends to non-homogeneous settings

$X_i^{(n)} = \frac{1}{n}\{\#\text{ objects in state } i\}$ $\Rightarrow X^{(n)}$ is not Markovian.

This extends to non-homogeneous settings

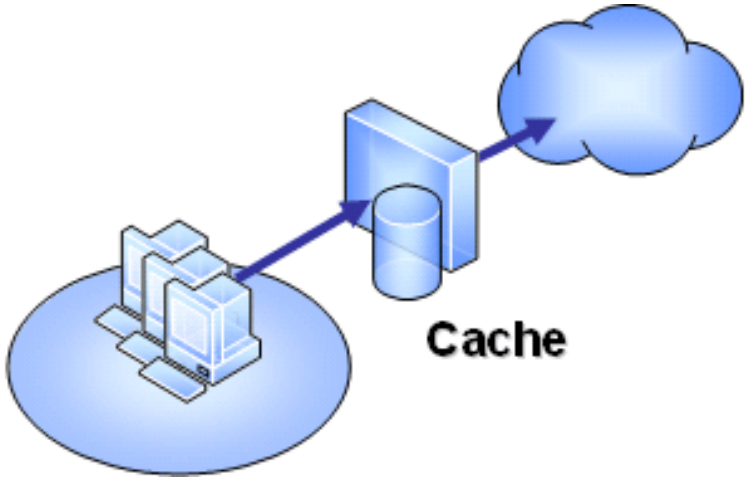
$\cancel{X_i^{(n)} = \frac{1}{n}\{\#\text{ objects in state } i\} \qquad \Rightarrow X^{(n)} \text{ is not Markovian.}}$

The solution is to use *one-hot encoding*:

$$Y_{(k,i)} = \begin{cases} 1 & \text{if object } k \text{ is in state } i \text{ at time } t \\ 0 & \text{otherwise} \end{cases}$$

$\mathbf{Y}^{(n)}$ is Markovian. We can construct a hydrodynamic limit for $\mathbf{y}$.

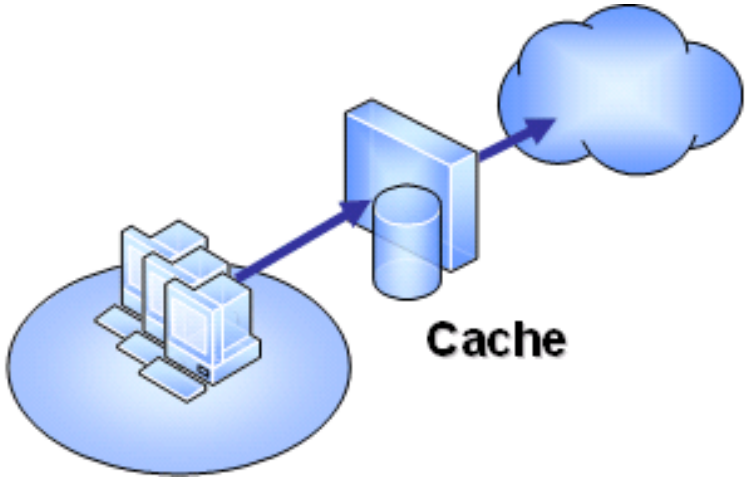
Example: cache replacement policies



Requests for k arrive at rate λ_k.

Random replacement policy.

Example: cache replacement policies



Requests for k arrive at rate λ_k.

Random replacement policy.

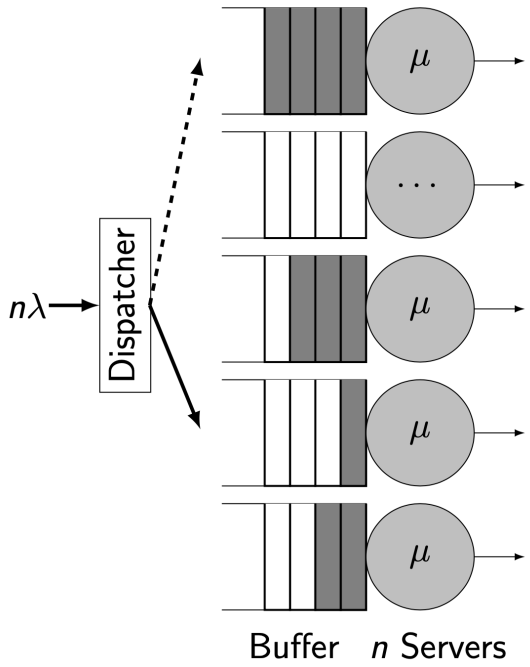
$Y_k = 1$ if object k is in the cache.

$$\dot{y}_k = \underbrace{\lambda_k(1-y_k)}_{\text{object } k \text{ is requested while outside}} - \underbrace{\sum_j \lambda_j(1-y_j)}_{\text{another object enters}} \underbrace{\frac{y_k}{\#\text{ cache size}}}_{\text{object } k \text{ is replaced}} .$$

Outline

Outline

Density dependent population process



Example: SQ(2) in the supermarket model

Mean Field Methodology:

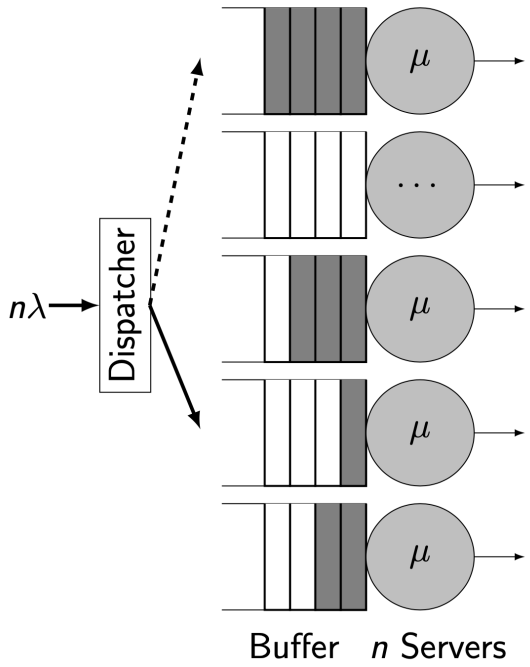
- $X_s^{(n)}(t) = \frac{1}{n}\{\#\text{ objects in state } s \text{ at } t\}$

Kurtz's density dependent population model:

$$X^{(n)} \to X^{(n)} + \frac{1}{n}\ell \qquad \text{at rate } n\beta_\ell(X)$$

Drift : $f(x) = \sum_\ell \ell\beta_\ell(x)$.

Density dependent population process



Example: SQ(2) in the supermarket model

Mean Field Methodology:

- $X_s^{(n)}(t) = \frac{1}{n}\{\#\text{ objects in state } s \text{ at } t\}$

Kurtz's density dependent population model:

$$X^{(n)} \to X^{(n)} + \frac{1}{n}\ell \qquad \text{at rate } n\beta_\ell(X)$$

Drift : $f(x) = \sum_\ell \ell\beta_\ell(x)$.

The mean field approximation is the solution of $\dot{x} = f(x)$

What is the bias of mean field approximaiton?

Consider a density dependent population process in $\mathbb{R}^d$ and assume that $\beta_\ell(x)$ are bounded.

Theorem (G., Bortolussi, Tribastone 2019) If the drift is C^2, there exists an (easily computable) vector $V(t)$ such that for any finite time:

$$\mathbb{E}[X_t] = \underbrace{\phi_t(X_0)}_{\text{mean field approx.}} + \underbrace{\frac{1}{n}V(t)}_{\text{First order bias}} + O(\frac{1}{n^2}).$$

This holds uniformly in time if the ODE has a unique exponentially stable attractor.

$V(t)$ is the first-order expansion of the bias of the approximation.

The expansion is in general very accurate for small n

n	10	100	1000	$+\infty$
Average queue length for SQ(2), $\rho = 0.9$	2.804	2.393	2.357	–

The expansion is in general very accurate for small n

n	10	100	1000	$+\infty$
Average queue length for SQ(2), $\rho = 0.9$	2.804	2.393	2.357	–
Refined approximation	2.751	2.393	2.357	2.353

where

- mean field $= \Phi_t(x)$.
- Refined = mean-field + V/n.

Intuition: Where does the $1/n$ term come from?

The moment closure approach

Consider a system for which X becomes $X + 1/n$ at rate nX^2. We have:

$$\frac{d}{dt}\mathbb{E}[X] = \mathbb{E}\left[X^2\right]$$

This equation is not closed

Intuition: Where does the $1/n$ term come from?

The moment closure approach

Consider a system for which X becomes $X + 1/n$ at rate nX^2. We have:

$$\frac{d}{dt}\mathbb{E}[X] = \mathbb{E}\left[X^2\right] \approx \mathbb{E}[X]^2 \text{ (mean field approx.)}$$

This equation is ~~not~~ closed

Intuition: Where does the $1/n$ term come from?

The moment closure approach

Consider a system for which X becomes $X + 1/n$ at rate nX^2. We have:

$$\frac{d}{dt}\mathbb{E}[X] = \mathbb{E}\left[X^2\right] \qquad \cancel{\approx \mathbb{E}[X]^2 \text{ (mean field approx.)}}$$

$$\frac{d}{dt}\mathbb{E}\left[X^2\right] = 2\mathbb{E}\left[X^3\right] + \frac{1}{n}\mathbb{E}\left[X^2\right]$$

This equation is not closed

Intuition: Where does the $1/n$ term come from?

The moment closure approach

Consider a system for which X becomes $X + 1/n$ at rate nX^2. We have:

$$\frac{d}{dt}\mathbb{E}[X] = \mathbb{E}\left[X^2\right] \qquad \cancel{\approx \mathbb{E}[X]^2 \text{ (mean field approx.)}}$$

$$\frac{d}{dt}\mathbb{E}\left[X^2\right] = 2\mathbb{E}\left[X^3\right] + \frac{1}{n}\mathbb{E}\left[X^2\right] \qquad \approx 2(3\mathbb{E}\left[X^2\right]\mathbb{E}[X] - 2\mathbb{E}[X]^2) + \frac{1}{n}\mathbb{E}\left[X^2\right]$$

(refined approximation)

This equation is ~~not~~ closed

Intuition: Where does the $1/n$ term come from?

The moment closure approach

Consider a system for which X becomes $X + 1/n$ at rate nX^2. We have:

$$\frac{d}{dt}\mathbb{E}[X] = \mathbb{E}\left[X^2\right] \qquad \cancel{\approx \mathbb{E}[X]^2 \text{ (mean field approx.)}}$$

$$\frac{d}{dt}\mathbb{E}\left[X^2\right] = 2\mathbb{E}\left[X^3\right] + \frac{1}{n}\mathbb{E}\left[X^2\right] \qquad \cancel{\approx 2(3\mathbb{E}\left[X^2\right]\mathbb{E}[X] - 2\mathbb{E}[X]^2) + \frac{1}{n}\mathbb{E}\left[X^2\right]}$$

$$\frac{d}{dt}\mathbb{E}\left[X^3\right] = \mathbb{E}\left[\frac{3X^4}{n} + \frac{4X^3}{n^2} + \frac{X^2}{n^3}\right]$$

$$\vdots$$

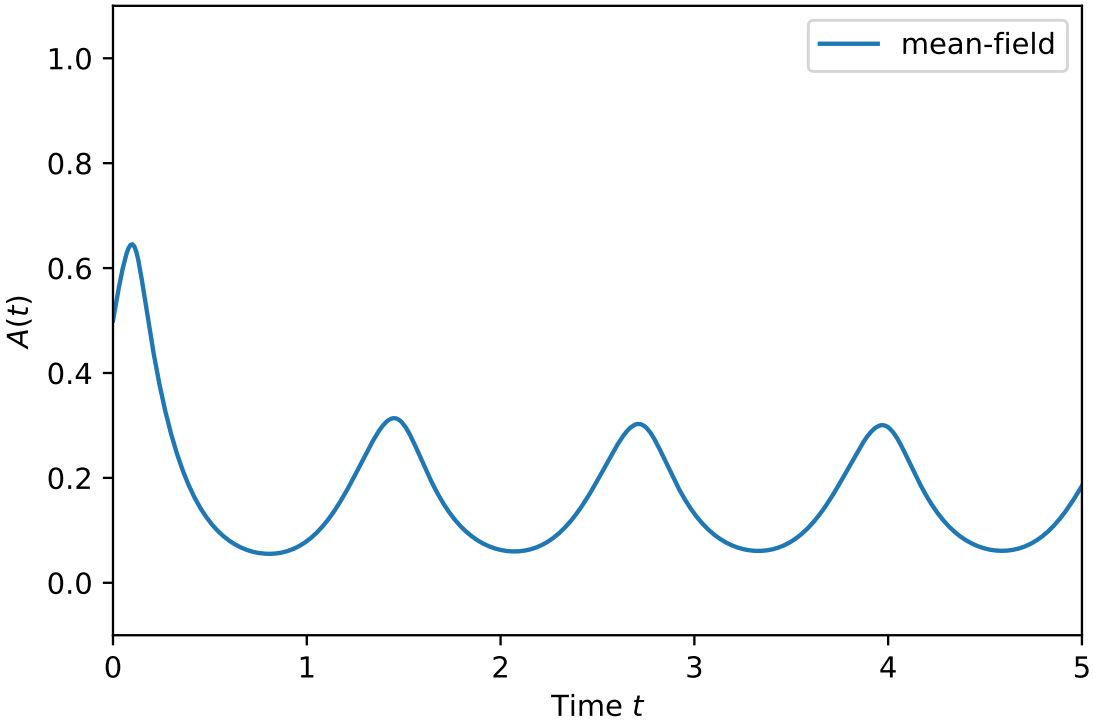
The moment equations are never closed.

- They can be closed by assuming $\mathbb{E}\left[(X - \mathbb{E}[X])^d\right] \approx 0$
- This gives a $O(1/n^{\lfloor (d+1)/2 \rfloor})$-accurate approximation.

Does it always work?

Does it always work?

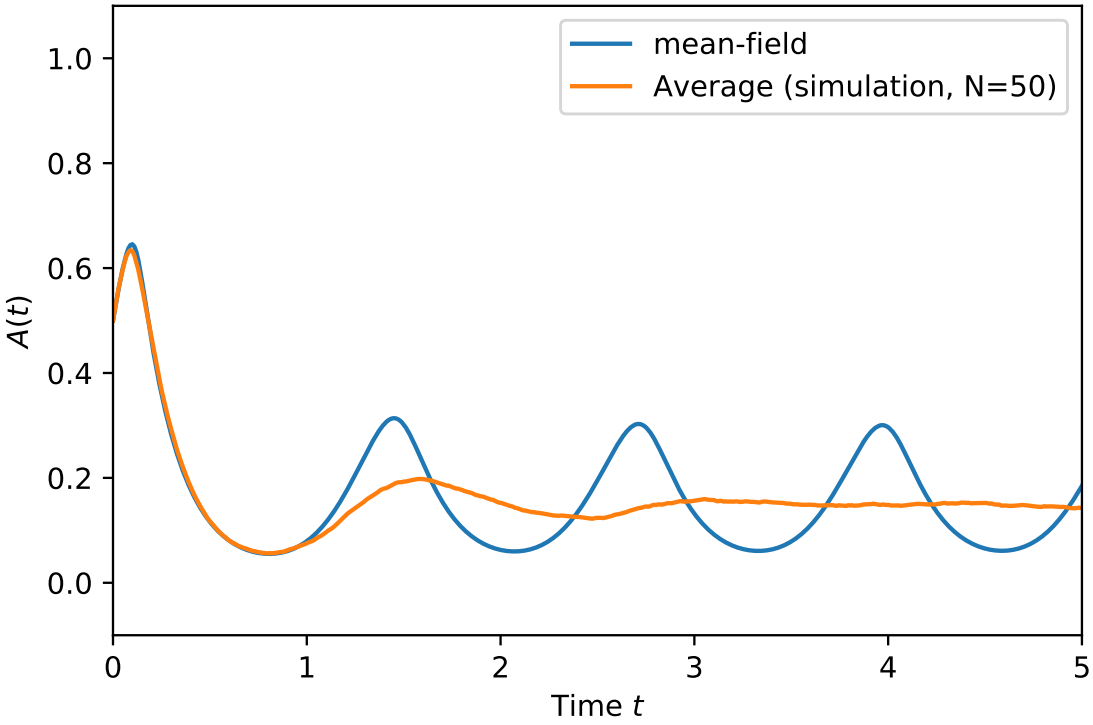
1. Interchange of limit : does $\lim_{N\to\infty} \lim_{t\to\infty} = \lim_{t\to\infty} \lim_{N\to\infty}$?



Example: SIR model with cyclic behavior.

Does it always work?

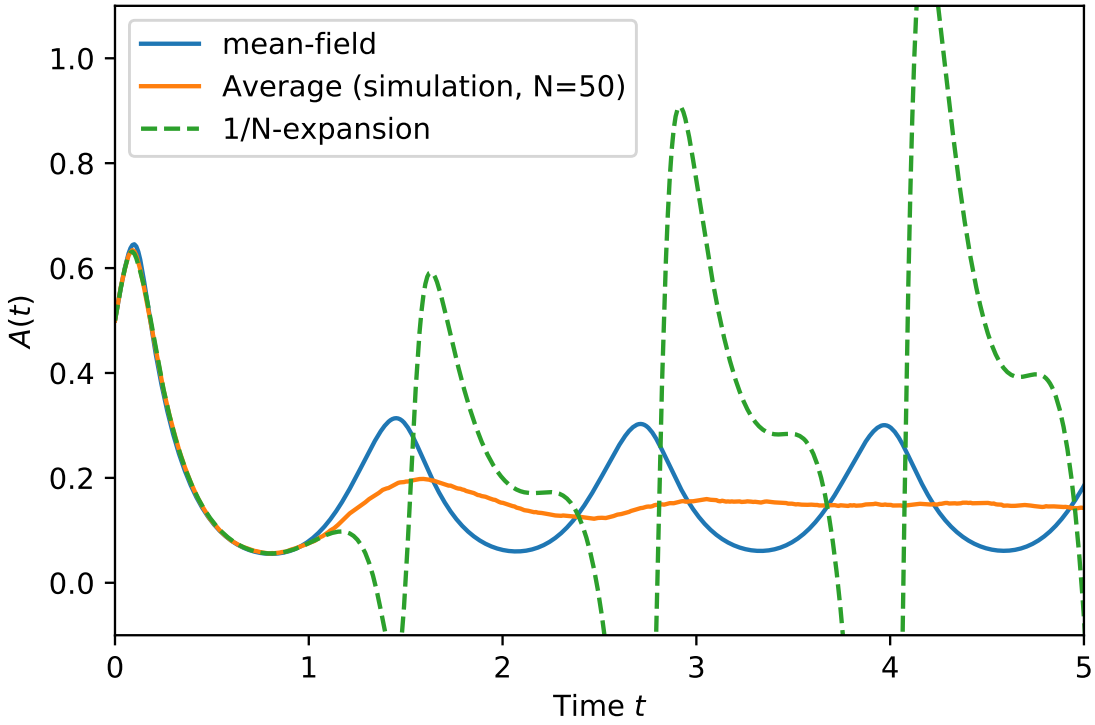
1. Interchange of limit : does $\lim_{N\to\infty} \lim_{t\to\infty} = \lim_{t\to\infty} \lim_{N\to\infty}$?



Example: SIR model with cyclic behavior.

Does it always work?

1. Interchange of limit : does $\lim_{N\to\infty} \lim_{t\to\infty} = \lim_{t\to\infty} \lim_{N\to\infty}$?



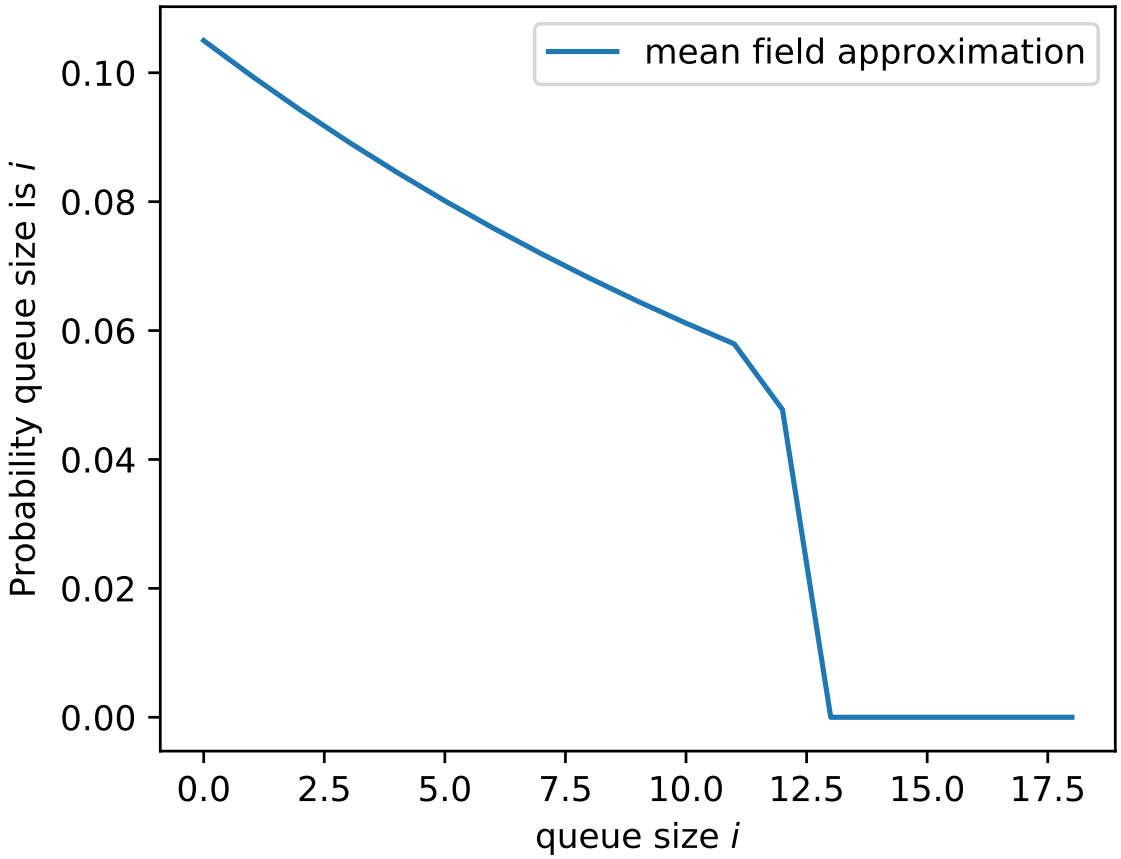
Example: SIR model with cyclic behavior.

Does it always work?

1. Interchange of limit

2. Non-smooth dynamics

Xu, Tsitsiklis 2011



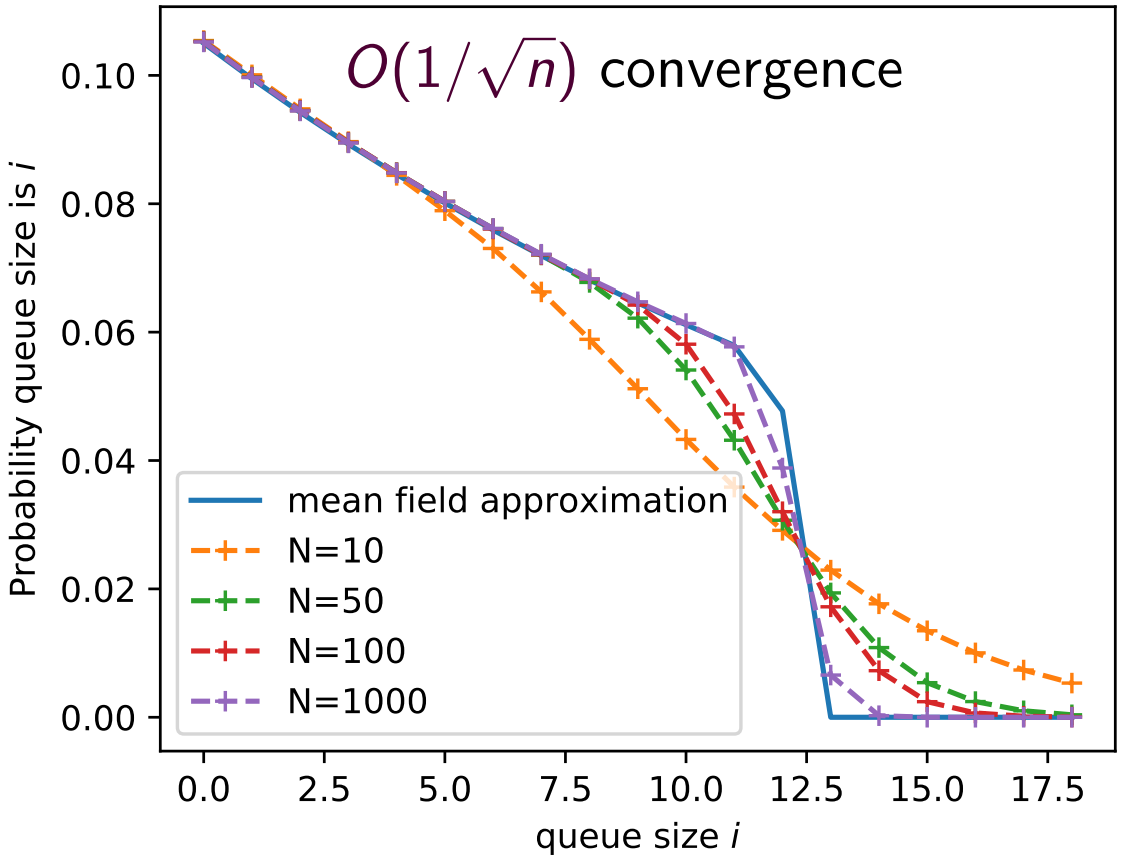
Power of (a little of) centralization

Does it always work?

1. Interchange of limit

2. Non-smooth dynamics

Xu, Tsitsiklis 2011



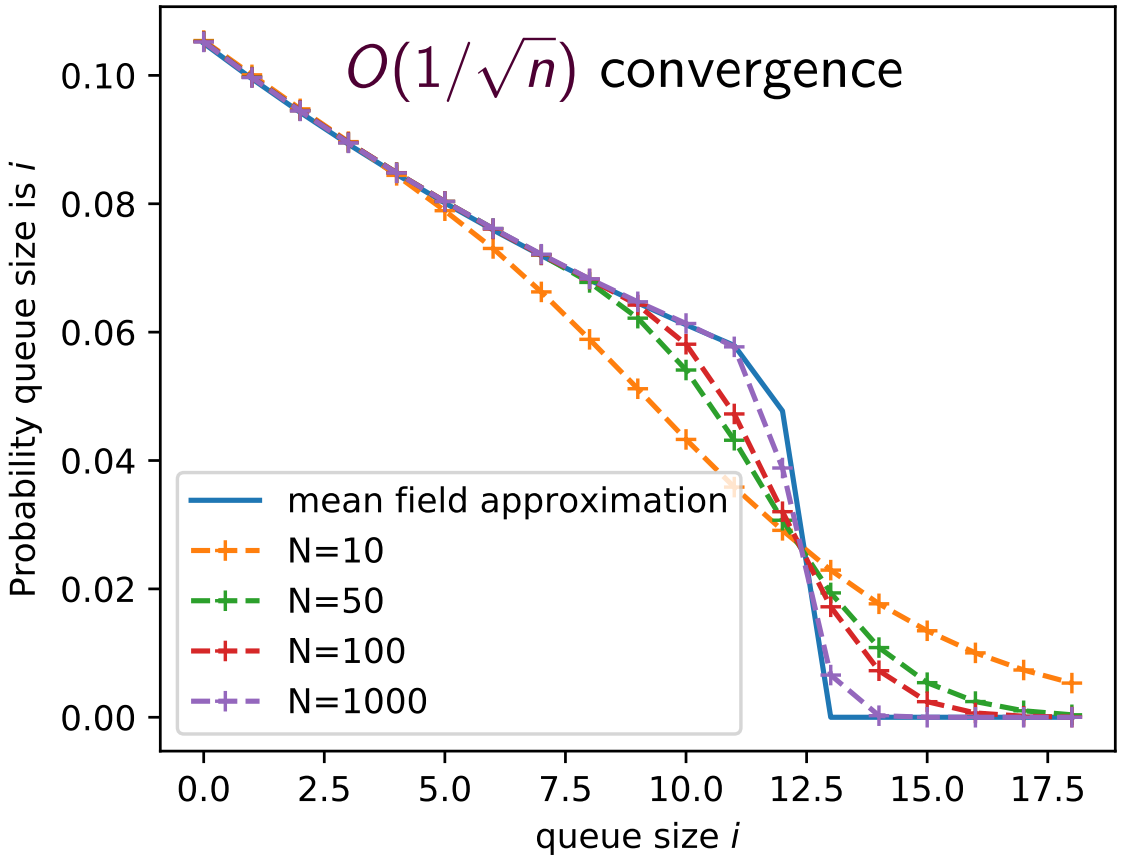
Power of (a little of) centralization

Does it always work?

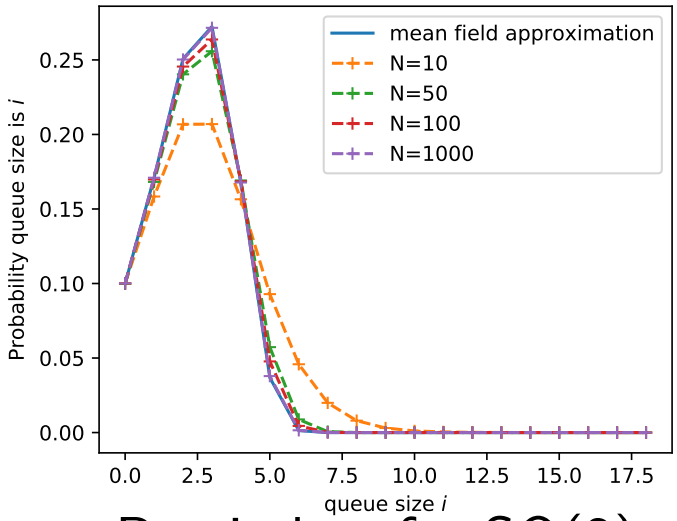
1. Interchange of limit

2. Non-smooth dynamics

Xu, Tsitsiklis 2011



Power of (a little of) centralization

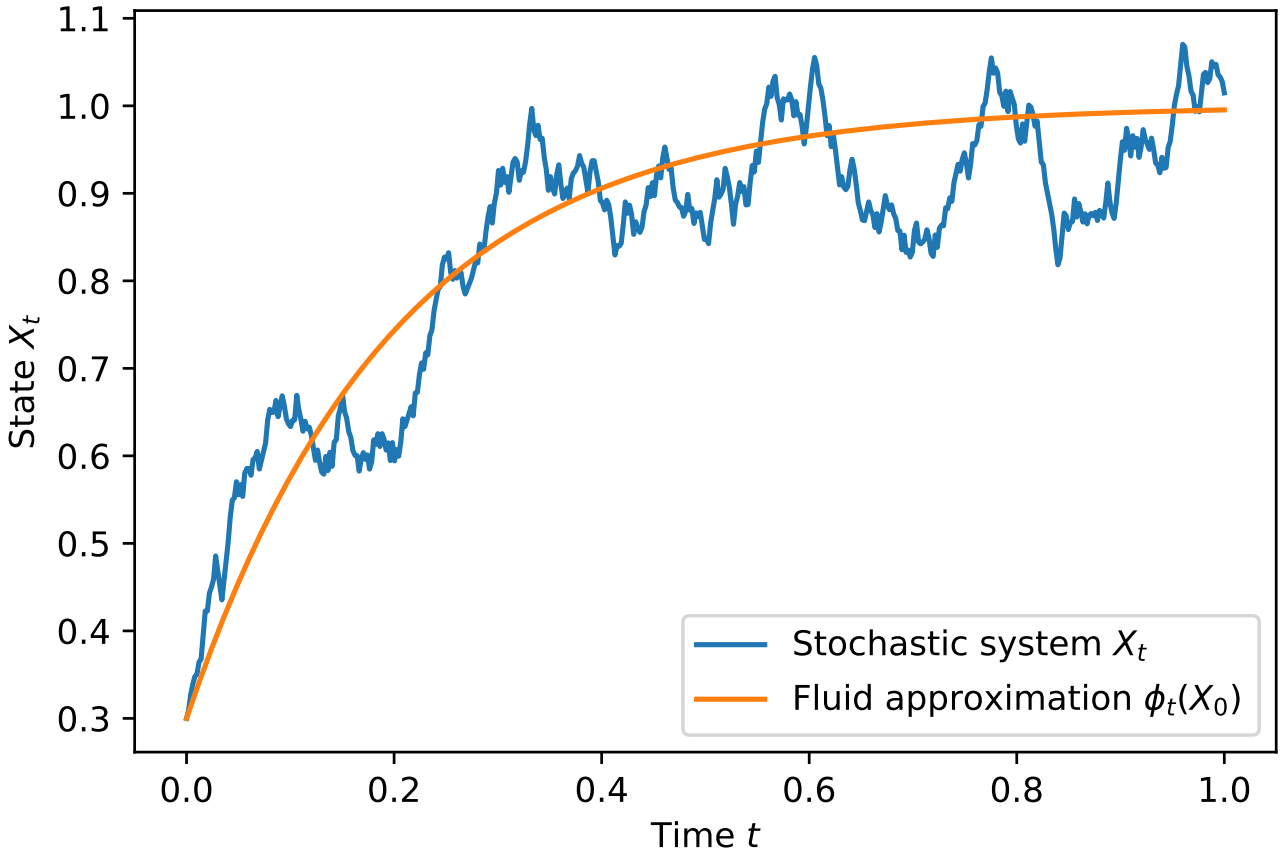


Reminder: for SQ(2)

$O(1/n)$ convergence

Outline

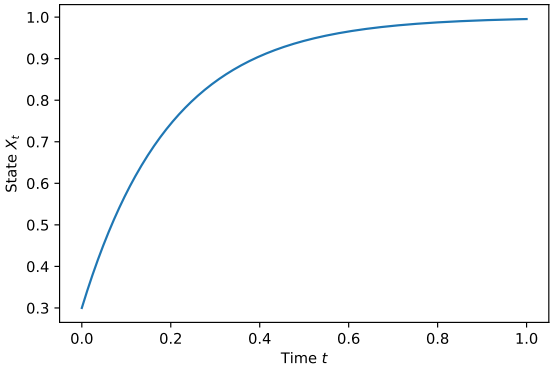
We compare a stochastic system and a fluid approximation



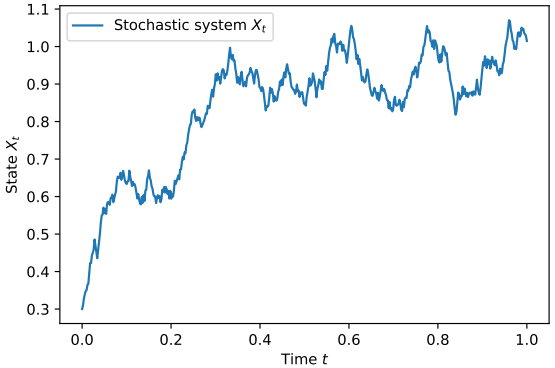
Important notations:

- Stochastic system $X_t \in \mathcal{X}$.
- Fluid approximation $\dot{x} = f(x)$. Solution starting from X_0 is $\phi_t(X_0)$.

To compare X_t and $\phi_t(X_0)$, we zoom on infinitesimal changes



$\Phi_t(X_0)$

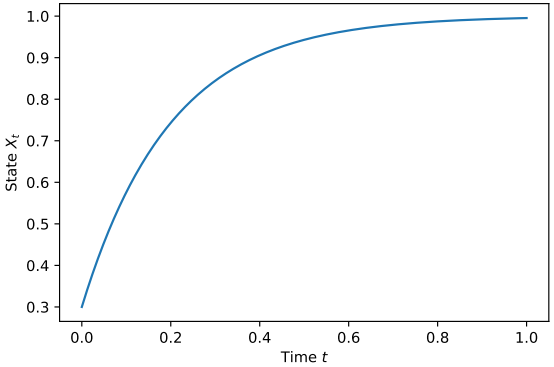


X_t

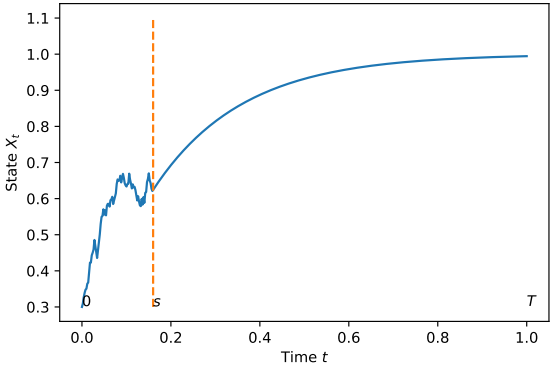
We want to compare:

$$\mathbb{E}\left[X_t\right] - \phi_t(X_0)$$

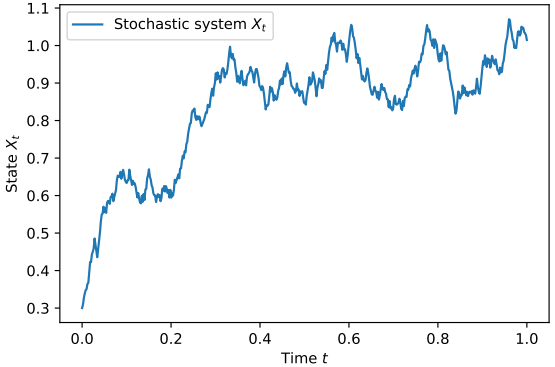
To compare X_t and $\phi_t(X_0)$, we zoom on infinitesimal changes



$\Phi_t(X_0)$



$\phi_{t-s}(X_s)$

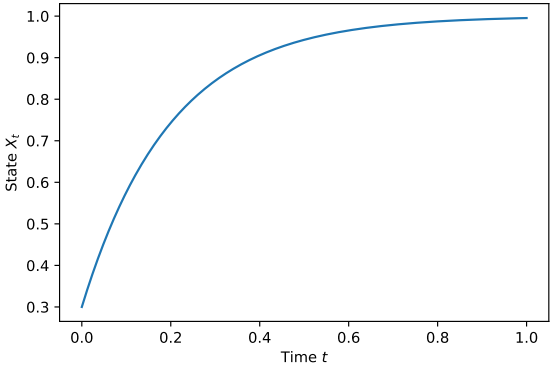


X_t

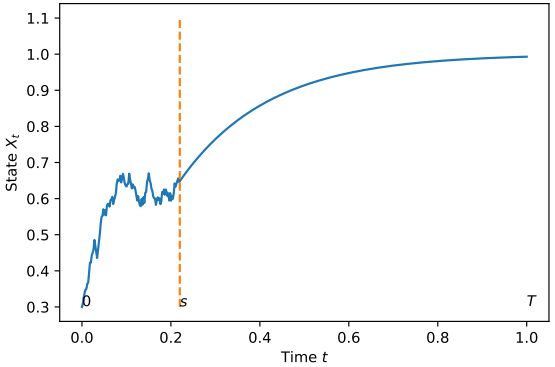
We want to compare:

$$\mathbb{E}\left[X_t\right] - \phi_t(X_0)$$

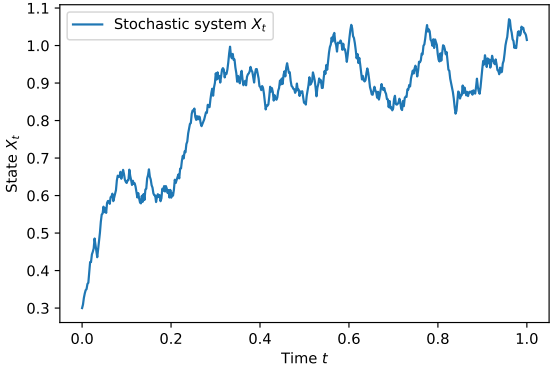
To compare X_t and $\phi_t(X_0)$, we zoom on infinitesimal changes



$\Phi_t(X_0)$



$\phi_{t-s}(X_s)$



X_t

We want to compare:

$$\mathbb{E}\left[X_t\right] - \phi_t(X_0) = \int_0^t \mathbb{E}\left[\frac{d}{ds}\phi_{t-s}(X_s)\right] ds$$

To study infinitesimal changes, we need generators

G^{sto} Generator of the stochastic system. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$G^{\mathrm{sto}} h(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}\left[h(X_t) - h(X_0) \mid X_0 = x\right].$$

Example: if X_t is a Markov chain of generator K:

$$G^{\mathrm{sto}} h(x) = \sum_{y \in \mathcal{X}} K_{xy}(h(y) - h(x)).$$

To study infinitesimal changes, we need generators

G^{sto} Generator of the stochastic system. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$G^{\mathrm{sto}} h(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}\left[h(X_t) - h(X_0) \mid X_0 = x\right].$$

G^{ODE} Generator of the ODE. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$\begin{aligned} G^{\mathrm{ODE}} h(x) &= \lim_{t \to 0} \frac{1}{t} (h(\Phi_t(x)) - h(x)) \\ &= \nabla h(x) \cdot f(x). \end{aligned}$$

Typically $f(x) = G^{\mathrm{sto}} I(x)$, where I is the identity function.

To study infinitesimal changes, we need generators

G^{sto} Generator of the stochastic system. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$G^{\mathrm{sto}} h(x) = \lim_{t \to 0} \frac{1}{t} \mathbb{E}\left[h(X_t) - h(X_0) \mid X_0 = x\right].$$

G^{ODE} Generator of the ODE. For a test function $h : \mathcal{X} \to \mathbb{R}$:

$$G^{\mathrm{ODE}} h(x) = \lim_{t \to 0} \frac{1}{t} (h(\Phi_t(x)) - h(x))$$

Using the generators, we can compare the two systems

$$\mathbb{E}\left[\frac{d}{ds}\phi_{t-s}(X_s)\right] ds$$

Using the generators, we can compare the two systems

$$\mathbb{E}\left[\frac{d}{ds}\phi_{t-s}(X_s)\right] ds = \mathbb{E}\left[G^{\mathrm{sto}}\phi_{t-s}(X_s) - G^{\mathrm{ODE}}\phi_{t-s}(X_s)\right]$$
$$= \mathbb{E}\left[(G^{\mathrm{sto}} - G^{\mathrm{ODE}})\phi_{t-s}(X_s)\right]$$

Using the generators, we can compare the two systems

$$\mathbb{E}\left[\frac{d}{ds}\phi_{t-s}(X_s)\right] ds = \mathbb{E}\left[G^{\mathrm{sto}}\phi_{t-s}(X_s) - G^{\mathrm{ODE}}\phi_{t-s}(X_s)\right]$$
$$= \mathbb{E}\left[(G^{\mathrm{sto}} - G^{\mathrm{ODE}})\phi_{t-s}(X_s)\right]$$

In particular:

$$\mathbb{E}\left[X_t\right] - \phi_t(X_0) = \int_0^t \mathbb{E}\left[(G^{\mathrm{sto}} - G^{\mathrm{ODE}})\phi_{t-s}(X_s)\right] ds$$
$$= (G^{\mathrm{sto}} - G^{\mathrm{ODE}}) \int_0^t \mathbb{E}\left[\phi_{t-s}(X_s)\right] ds$$

(Taking the limit $t \to \infty$, we obtain Stein's method).

Recap

For finite t:

$$\underbrace{\mathbb{E}\left[X_t\right]}_{\text{Stochastic system}} - \underbrace{\phi_t(X_0)}_{\text{deterministic approx.}} = (G^{\text{sto}} - G^{\text{ODE}}) \int_0^t \mathbb{E}\left[\phi_{t-s}(X_s)\right] ds$$

For $t = +\infty$, if $x^* = \phi_\infty(X_0)$ does not depend on X_0, we have:

$$\underbrace{\mathbb{E}\left[h(X_\infty)\right]}_{\text{Stochastic system}} - \underbrace{h(x^*)}_{\text{deterministic approx.}} = (G^{\text{sto}} - G^{\text{ODE}}) \int_0^\infty \mathbb{E}\left[\phi_t(X_\infty)\right] ds.$$

Recap

For finite t:

$$\underbrace{\mathbb{E}\left[X_t\right]}_{\text{Stochastic system}} - \underbrace{\phi_t(X_0)}_{\text{deterministic approx.}} = (G^{\text{sto}} - G^{\text{ODE}}) \int_0^t \mathbb{E}\left[\phi_{t-s}(X_s)\right] ds$$

For $t = +\infty$, if $x^* = \phi_\infty(X_0)$ does not depend on X_0, we have:

$$\underbrace{\mathbb{E}\left[h(X_\infty)\right]}_{\text{Stochastic system}} - \underbrace{h(x^*)}_{\text{deterministic approx.}} = (G^{\text{sto}} - G^{\text{ODE}}) \int_0^\infty \mathbb{E}\left[\phi_t(X_\infty)\right] ds.$$

To prove that the sto $\approx$ deterministic, we prove that:

- for some $h \in \mathcal{H}$, $(G^{\text{sto}} - G^{\text{ODE}})h$ is small.
- $\int_0^t \mathbb{E}\left[\phi_{t-s}\right]$ or $\int_0^\infty \mathbb{E}\left[\phi_t\right]$ belongs to this $\mathcal{H}$.

Application to density dependent population processes (1/2)

The generator of the density dependent population process is:

$$G^{\mathrm{sto}}h(x) = \sum_{\ell} \left(h(x + \frac{1}{n}\ell) - h(x) \right) n\beta_\ell(x)$$

$$= \underbrace{\nabla h \cdot f(x)}_{\text{generator of the ODE, } G^{\mathrm{ODE}}} + \frac{1}{n} \underbrace{\nabla^2 h \cdot Q(x)}_{\text{Diffusion term}} + O(1/n^2).$$

Application to density dependent population processes (1/2)

The generator of the density dependent population process is:

$$G^{\mathrm{sto}}h(x) = \sum_{\ell} \left(h(x + \frac{1}{n}\ell) - h(x) \right) n\beta_\ell(x)$$

$$= \underbrace{\nabla h \cdot f(x)}_{\text{generator of the ODE, } G^{\mathrm{ODE}}} + \frac{1}{n} \underbrace{\nabla^2 h \cdot Q(x)}_{\text{Diffusion term}} + O(1/n^2).$$

As a consequence

- $(G^{\mathrm{sto}} - G^{\mathrm{ODE}})h = O(1/n)$ if h is C^2. $\Rightarrow$ Set $\mathcal{H}$ of slide 24 is C^2.
- The hidden constant depends on $\left\|\nabla^2 h \cdot Q\right\|$. Studying this gives $V(t)$.

Consequence for the error of mean field model (2/2)

- For finite-horizon, the function $h(x) = \int_0^t \phi_s(x) ds$ is C^2 if the drift function f is C^2.

- For infinite-horizon model, $h(x) = \int_0^\infty (\phi_s(x) - x^*) ds$ is C^2 if in addition x^* is an exponentially stable attractor.

The two functions belongs to "$\mathcal{H}$" $\Rightarrow$ Error $= O(1/n)$.

Some historical remarks

- Ying 2016: L_2 error is $O(1/\sqrt{n})$ for steady-state.
- G. 2017: Bias is $O(1/n)$.
- G. 2018, 2019: Expansion terms for the bias.

- G. Allmeier 2022: Extension to heterogeneous models.

More recently:

- This works for heterogeneous models.
- Error bounds for averaging methods (multi-scale models)

Outline

Conclusion

Mean field or fluid approximations are widely used heuristic.

- They simplify the analysis of stochastic systems.

We question its validity / accuracy.

Conclusion

Mean field or fluid approximations are widely used heuristic.

- They simplify the analysis of stochastic systems.

We question its validity / accuracy.

- We characterize the bias for different models (smooth homogeneous, heterogeneous, multi-scale).
- To do so, we take correlations into account.
- Numerical library: `https://pypi.org/project/rmftool/`

Conclusion

Mean field or fluid approximations are widely used heuristic.

- They simplify the analysis of stochastic systems.

We question its validity / accuracy.

- We characterize the bias for different models (smooth homogeneous, heterogeneous, multi-scale).
- To do so, we take correlations into account.
- Numerical library: `https://pypi.org/project/rmftool/`

Many open questions: optimization (bandit problems), (sparse) geometric models, non-Markovian.

Slides and references: `http://polaris.imag.fr/nicolas.gast`

References

Results on which this talk is based:

- Mean Field and Refined Mean Field Approximations for Heterogeneous Systems: It Works! by Allmeier and Gast. SIGMETRICS 2022.
- 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. Performance 2018.
- Two-scale: Bias and Refinement of Multiscale Mean Field Models. Allmeier, Gast, 2022 (arxiv).
  - CSMA model from CSMA networks in a many-sources regime: A mean-field approach. Cecchi, Borst, van Leeuwaarden, Whiting. Infocom 2016.

Paper cited as open problems:

- Pair-approximation: The Power of Two Choices on Graphs: the Pair-Approximation is Accurate by Gast. Mama 2015.
- Non-Markovian: Randomized Load Balancing with General Service Time Distributions by Bramson, Ly and Prabhakar. Sigmetrics 2010 and The PDE Method for the Analysis of Randomized Load Balancing Networks by Aghajani, Li, Ramanan.SIGMETRICS 2018