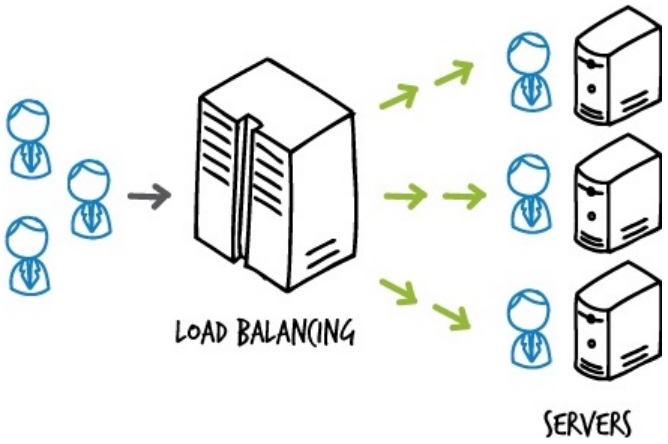
The Bias of Mean Field Approximation

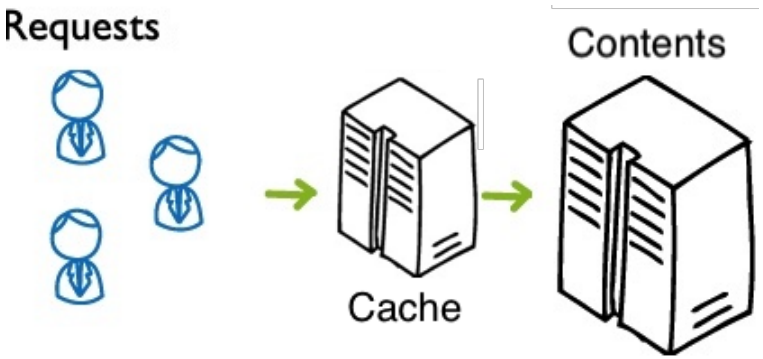
Nicolas Gast

Inria

Stochastic Networks Conference, June 2022. Cornell

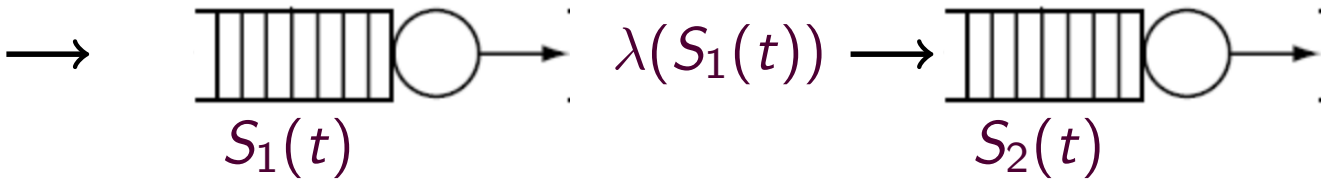
Stochastic networks = networks of stochastic objects





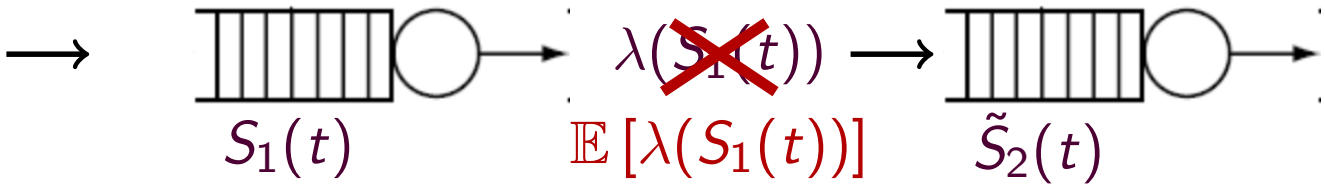
In this talk: Methods for quantitative evaluation

The curse of dimensionality



- $S_1(t)$ and $S_2(t)$ are dependent. $\Rightarrow n$ queues $= S^n$ possible states.

The curse of dimensionality disappears when using mean field approx.



- ~~$S_1(t)$ and $S_2(t)$ are dependent. $\Rightarrow$ n queues $=$ S^n possible states.~~
- $\tilde{S}_2(t)$ depends on $\mathbf{P}[S_1(t) = s]$ $\Rightarrow$ n objects $=$ Sn values.

Mean field assumption is to approximate $S_2(t) \sim \tilde{S}_2(t)$.

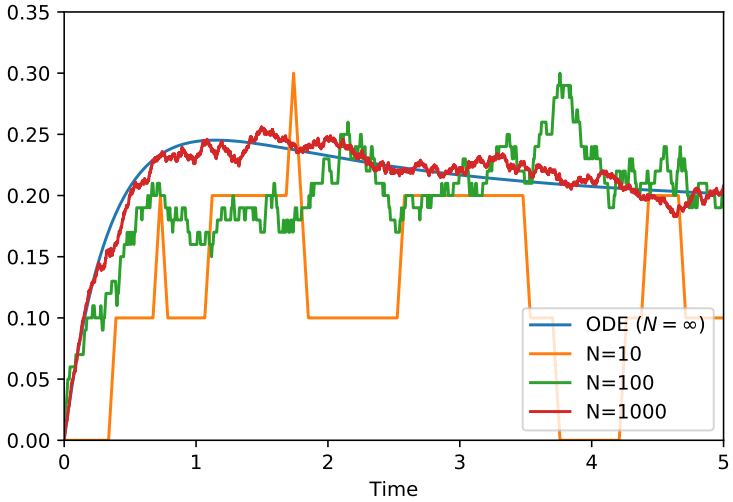
Mean field approx $\approx$ independence assumption.

How accurate is mean field approximation?

Theorem (Folk)

Mean field approximation is exact as n goes to infinity.

Proof: Law of large number.



Mean field-like methods are widely used

- Epidemiology (SIR-like models) (too many papers!)
- Cache replacement policy (TTL-approximation) Che et al 02, Fagin 77
- Replica mean field (neuroscience) Baccelli Davydov 22
- Bandit (Whittle-index) Whittle 88, Verloop 16
- Wireless (CSMA) Bianchi 00, Borst-Cecchi 16
- ...

Mean field-like methods are widely used

- Epidemiology (SIR-like models) (too many papers!)
- Cache replacement policy (TTL-approximation) Che et al 02, Fagin 77
- Replica mean field (neuroscience) Baccelli Davydov 22
- Bandit (Whittle-index) Whittle 88, Verloop 16
- Wireless (CSMA) Bianchi 00, Borst-Cecchi 16
- ...

Questions in this talk:

1. Why is this approximation so popular?
   - How accurate is it?
2. Can we do better approximation?

Outline

Outline

Interaction model

We consider a population of n objects with two types of interactions:

- Unilateral transitions:

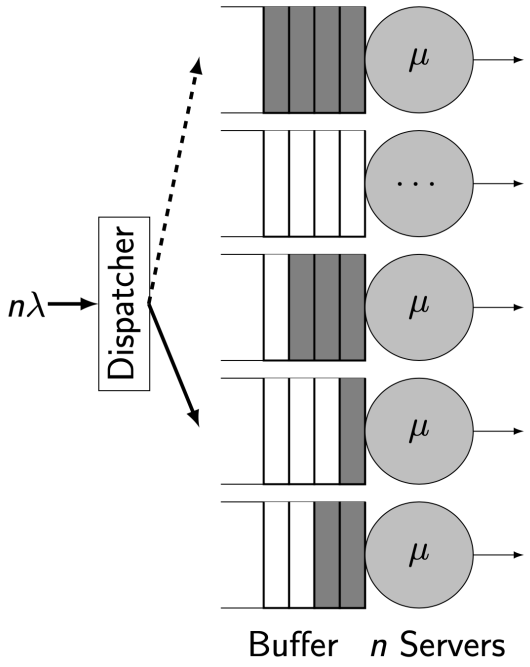
  Object k jumps from state i to j at rate $r_{ij}^{(k)}$

- Pairwise interactions:

  Object k, k' *simultaneously* jump from states (i, i') to (j, j') at rate $r_{ij,i'j'}^{(k,k')}/n$

If the rates do not depend on k, we call the model homogeneous.

Homogeneous: Classical Mean Field Setting



Example: Load-balancing

Mean Field Methodology:

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

Kurtz's density dependent population model:

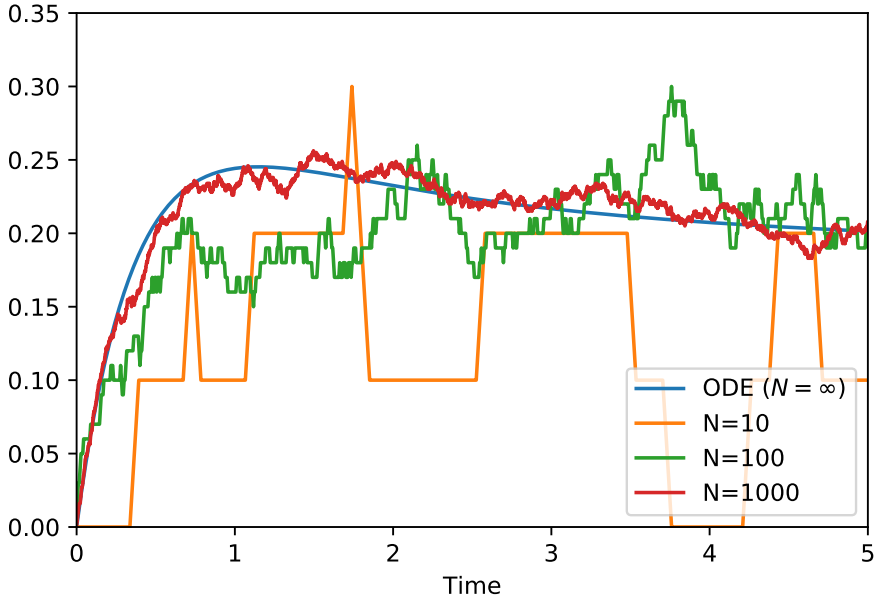
$$M^{(n)} \to M^{(n)} + \frac{1}{n}\ell \qquad \text{at rate } n\beta_\ell(M)$$

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

Accuracy of the Mean Field Approximation

For a system with n homogeneous objects with smooth drift f (C^1):

$$M^{(n)}(t) \approx \underbrace{m(t)}_{ODE\,=\,\text{mf approx}} + \underbrace{\frac{1}{\sqrt{n}} B_t}_{\text{noise}}.$$

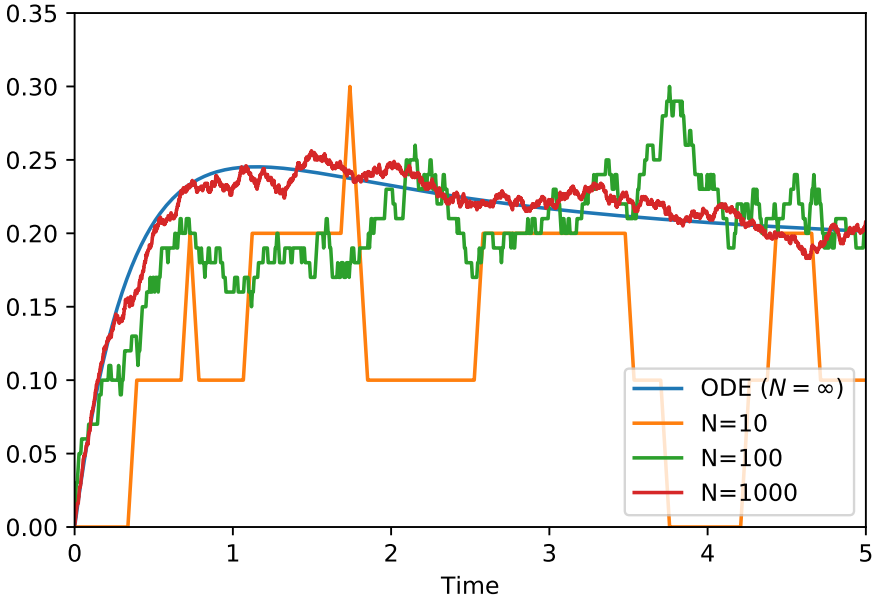


This result is about trajectories.

Accuracy of the Mean Field Approximation

For a system with n homogeneous objects with smooth drift f (C^1):

$$M^{(n)}(t) \approx \underbrace{m(t)}_{ODE = \text{mf approx}} + \underbrace{\frac{1}{\sqrt{n}} B_t}_{\text{noise}}.$$



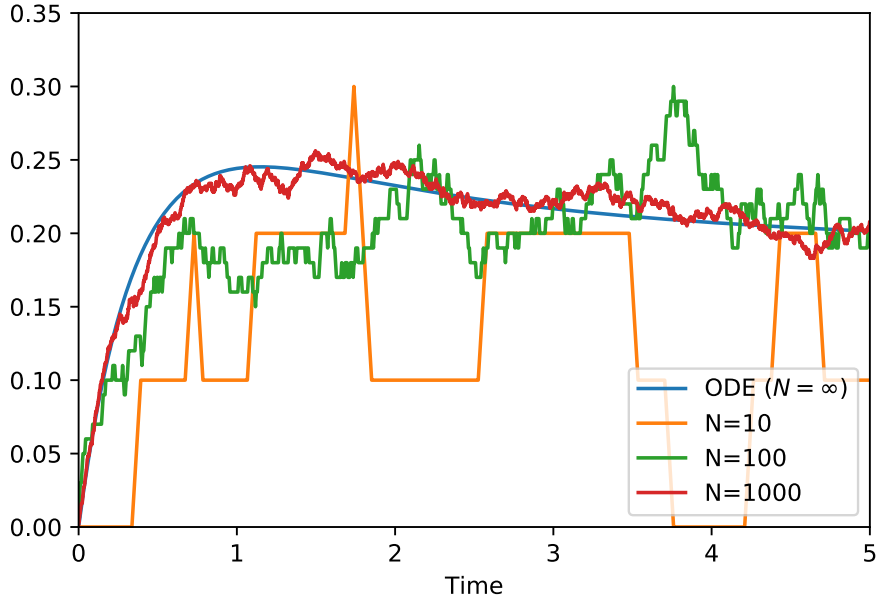
This result is about trajectories.

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

Accuracy of the Mean Field Approximation

For a system with n homogeneous objects with smooth drift f (C^1):

$$M^{(n)}(t) \approx \underbrace{m(t)}_{ODE = \text{mf approx}} + \underbrace{\frac{1}{\sqrt{n}} B_t}_{\text{noise}}.$$



This result is about trajectories.

n	10	100	1000	$+\infty$
Average queue length for SQ(2), $\rho = 0.9$	2.804	2.393	2.357	2.353
Bias	0.45	0.039	0.004	0

If the drift is C^2, then the bias $\mathbb{E}[M] - m$ is $O(1/n)$.

Theorem (Kolokoltsov 2012, G. 2017, G. and Van Houdt 2018)

For a DDPP, if the drift f is C^2, then for any[a] $t \in (0, \infty)$:

1. *There exists a (deterministic) vector $V(t)$ such that:*

$$\mathbb{E}\left[M^{(n)}(t)\right] = \underbrace{m(t)}_{\text{mean field approx.}} + \frac{V(t)}{n} + O(1/n^2)$$

2. *$V(t)$ can be (easily) computed numerically*

[a] Also holds for $t = +\infty$ if the ODE has an exponentially stable attractor.

If the drift is C^2, then the bias $\mathbb{E}[M] - m$ is $O(1/n)$.

We obtain a refined approximation

Theorem (Kolokoltsov 2012, G. 2017, G. and Van Houdt 2018)

For a DDPP, if the drift f is C^2, then for any[a] $t \in (0, \infty)$:

1. *There exists a (deterministic) vector $V(t)$ such that:*

$$\mathbb{E}\left[M^{(n)}(t)\right] = \underbrace{m(t) + \frac{V(t)}{n}}_{\textit{refined approximation}} + O(1/n^2)$$

2. *$V(t)$ can be (easily) computed numerically*

[a] Also holds for $t = +\infty$ if the ODE has an exponentially stable attractor.

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

Where does the $1/n$ comes from?

Two ingredients:

- Stein's method / comparison of generators.
- Moment closure approach.

Where does the $1/n$ comes from?

Two ingredients:

- Stein's method / comparison of generators.
- Moment closure approach.

Generators of the stochastic system: for a C^2 function h:

$$L^{(n)}h(m) = \sum_{\ell}(h(m + \frac{\ell}{n}) - h(m))n\beta_\ell(m)$$

Where does the $1/n$ comes from?

Two ingredients:

- Stein's method / comparison of generators.
- Moment closure approach.

Generators of the stochastic system: for a C^2 function h:

$$L^{(n)}h(m) = \sum_{\ell}(h(m+\frac{\ell}{n}) - h(m))n\beta_\ell(m)$$
$$= Dh(m)\cdot\underbrace{\sum_{\ell}\ell\beta_\ell(m)}_{=:f(m)} + \frac{1}{n}D^2h(m)\cdot\underbrace{\sum_{\ell}\ell\otimes\ell\beta_\ell(m)}_{Q(m)} + O(\frac{1}{n^2})$$

Where does the $1/n$ comes from?

Two ingredients:

- Stein's method / comparison of generators.
- Moment closure approach.

Generators of the stochastic system: for a C^2 function h:

$$
\begin{aligned}
L^{(n)}h(m) &= \sum_{\ell}(h(m+\frac{\ell}{n}) - h(m))n\beta_\ell(m) \\
&= Dh(m)\cdot\underbrace{\sum_{\ell}\ell\beta_\ell(m)}_{=:f(m)} + \frac{1}{n}D^2h(m)\cdot\underbrace{\sum_{\ell}\ell\otimes\ell\beta_\ell(m)}_{Q(m)} + O(\frac{1}{n^2}) \\
&= \underbrace{Dh(m)\cdot f(m)}_{\text{ODE generator (drift)}} + \frac{1}{n}\underbrace{D^2h(m)\cdot Q(m)}_{\text{Noise (covariance)}} + O(\frac{1}{n^2})
\end{aligned}
$$

Generator approach (continued)

If $\phi_t(m)$ is the solution of the ODE starting from m, then:

$$\begin{aligned}
\mathbb{E}\left[M(t) - \phi_t(M(0))\right] &= \int_0^t \frac{d}{ds}\mathbb{E}\left[\Phi_s(M^{(n)}(t-s))\right] ds \\
&= \int_0^t \mathbb{E}\left[(L - L^{(n)})\Phi_s(M^{(n)}(t-s))\right] ds \\
&= O(1/n) \qquad \text{if } \Phi_s \text{ is } C^2.
\end{aligned}$$

How to make it computable?

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

How to make it computable?

The moment closure approach

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

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

This equation is ~~not~~ closed

How to make it computable?

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

How to make it computable?

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

How to make it computable?

The moment closure approach

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

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

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

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

$$\vdots$$

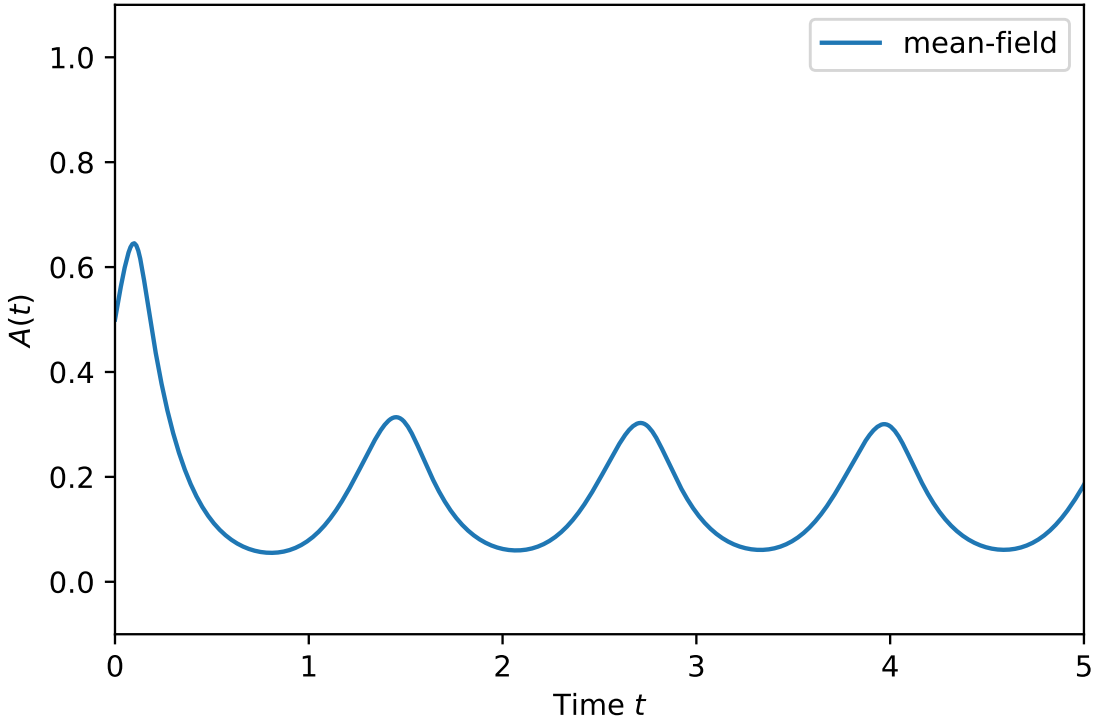
The moment equations are never closed.

- 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?

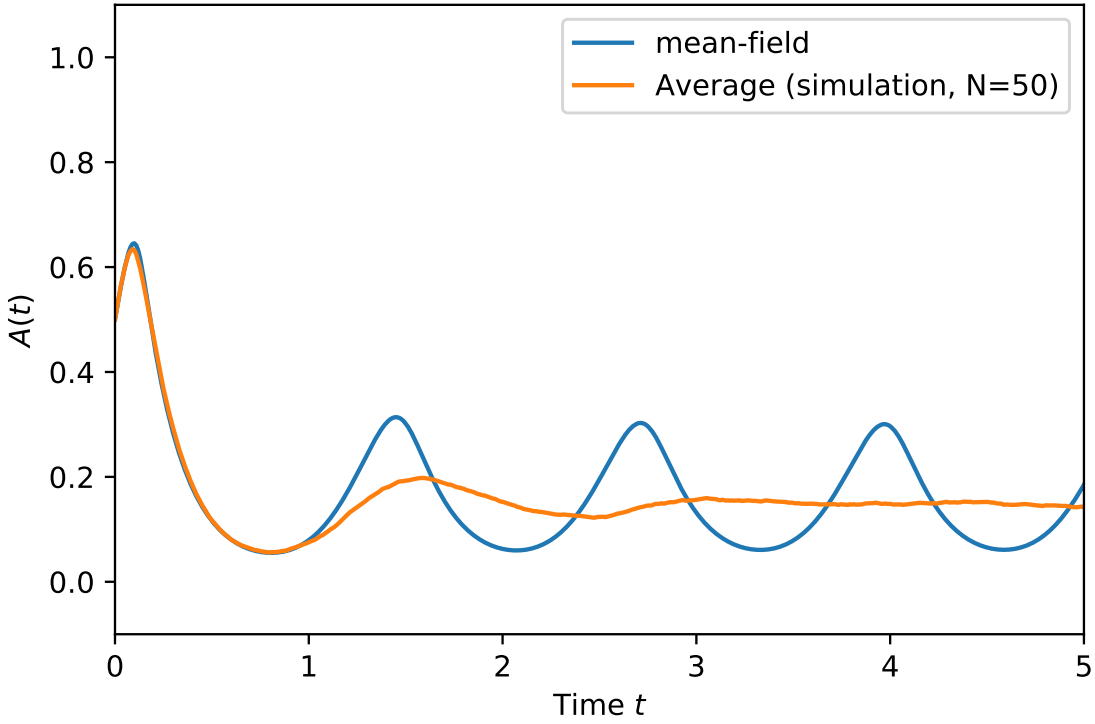
1. Interchange of limit : does $\lim_{N\to\infty} \lim_{t\to\infty} = \lim_{t\to\infty} \lim_{N\to\infty}$?



Example: SIR model with cyclic behavior.

Does it always work?

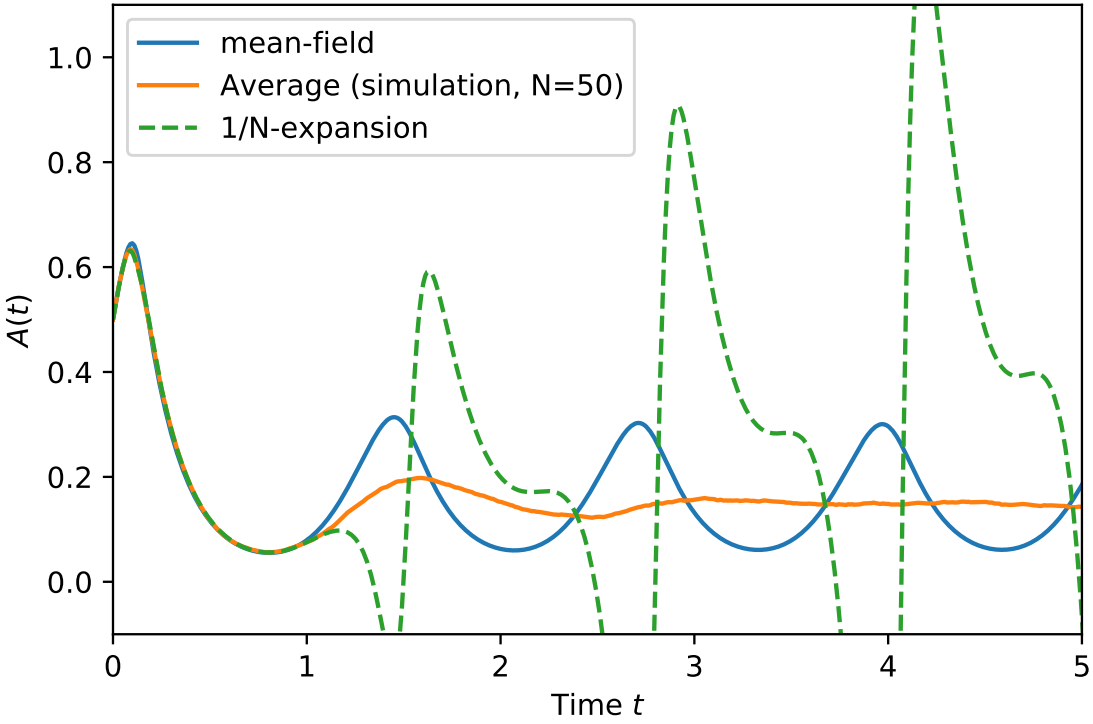
1. Interchange of limit : does $\lim_{N\to\infty} \lim_{t\to\infty} = \lim_{t\to\infty} \lim_{N\to\infty}$?



Example: SIR model with cyclic behavior.

Does it always work?

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

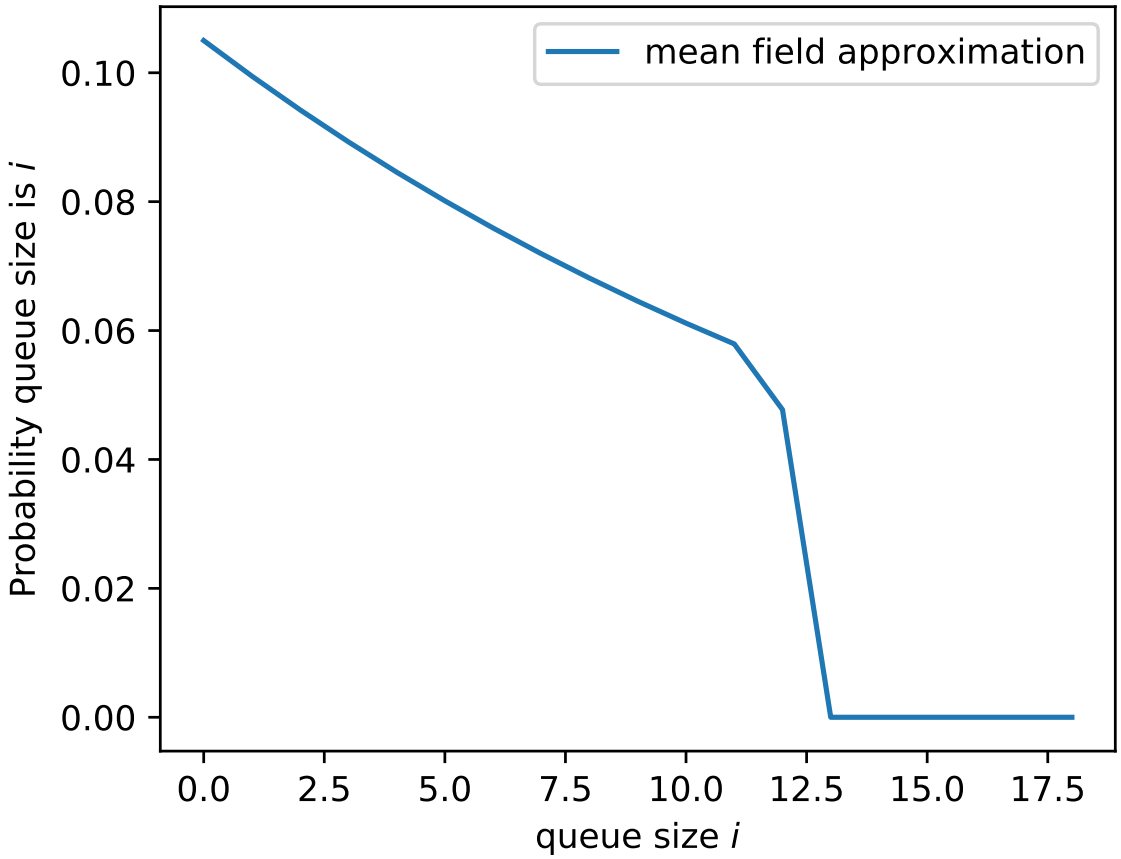


Example: SIR model with cyclic behavior.

Does it always work?

1. Interchange of limit
2. Non-smooth dynamics

Xu, Tsitsiklis 2011



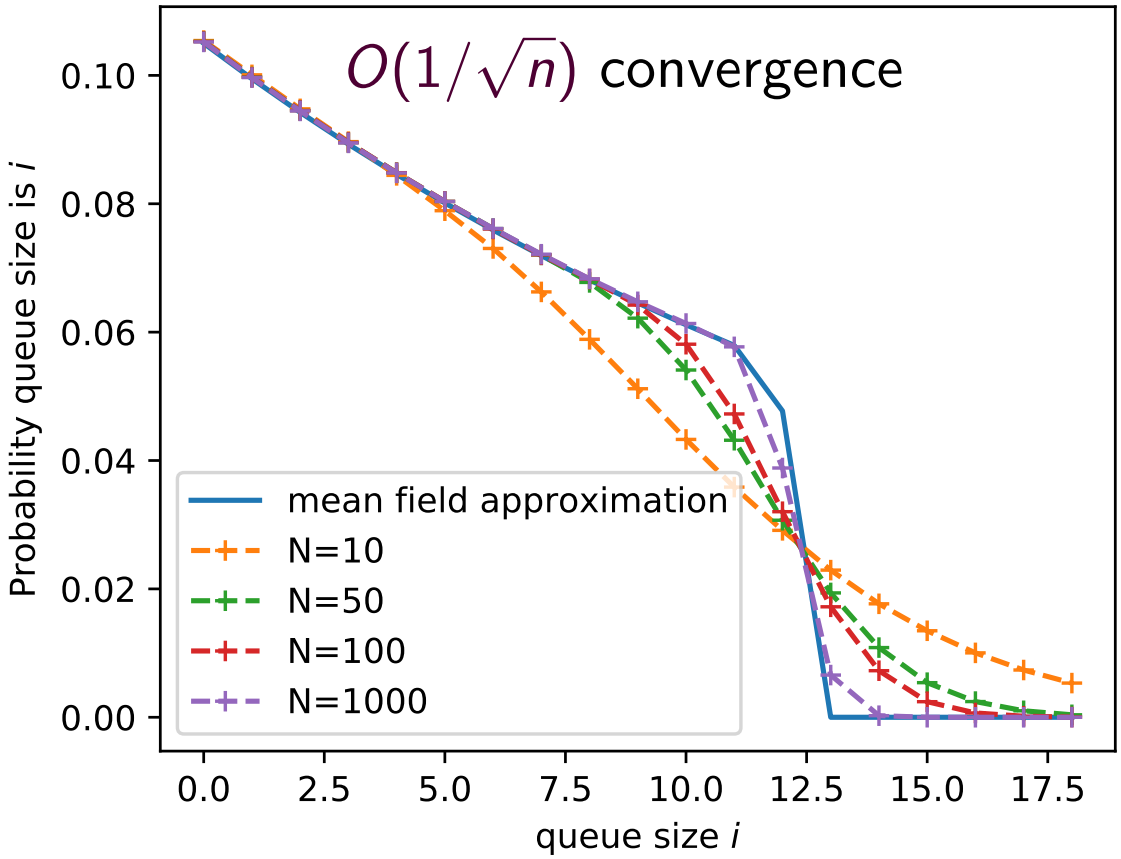
Power of (a little of) centralization

Does it always work?

1. Interchange of limit

2. Non-smooth dynamics

Xu, Tsitsiklis 2011



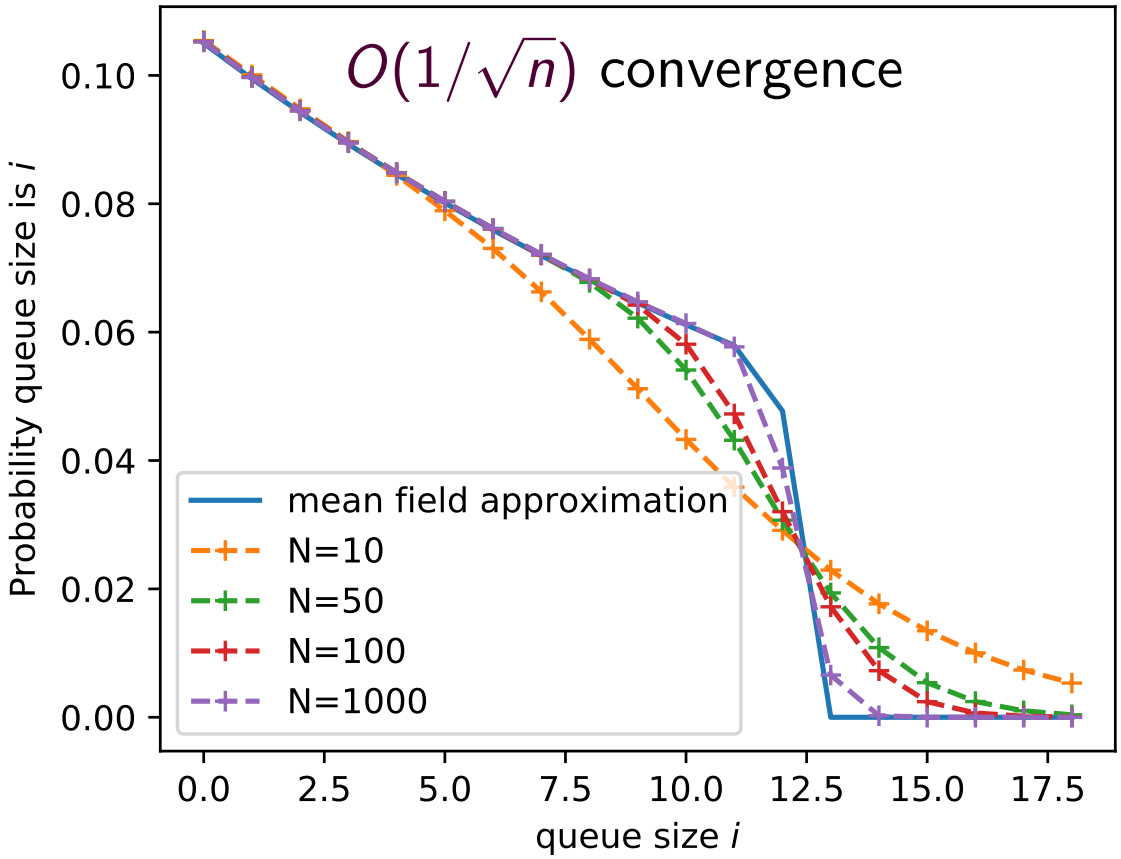
Power of (a little of) centralization

Does it always work?

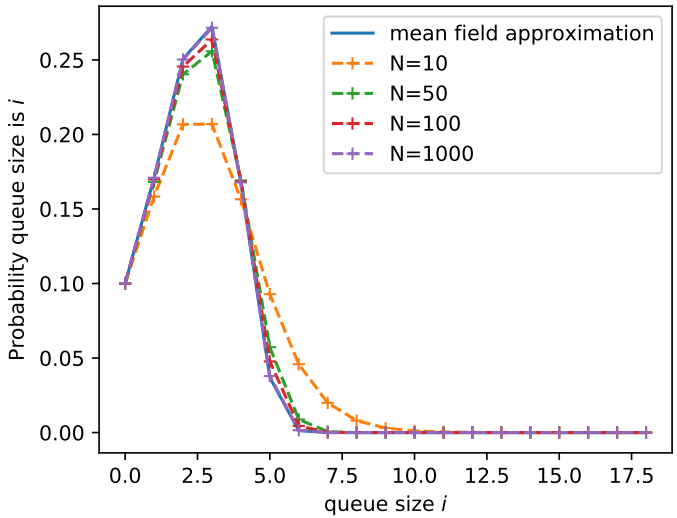
1. Interchange of limit

2. Non-smooth dynamics

Xu, Tsitsiklis 2011



Power of (a little of) centralization



Power two choice

$O(1/n)$ convergence

Recap on homoegeneous models

Mean field is often justified by studying pointwise convergence:

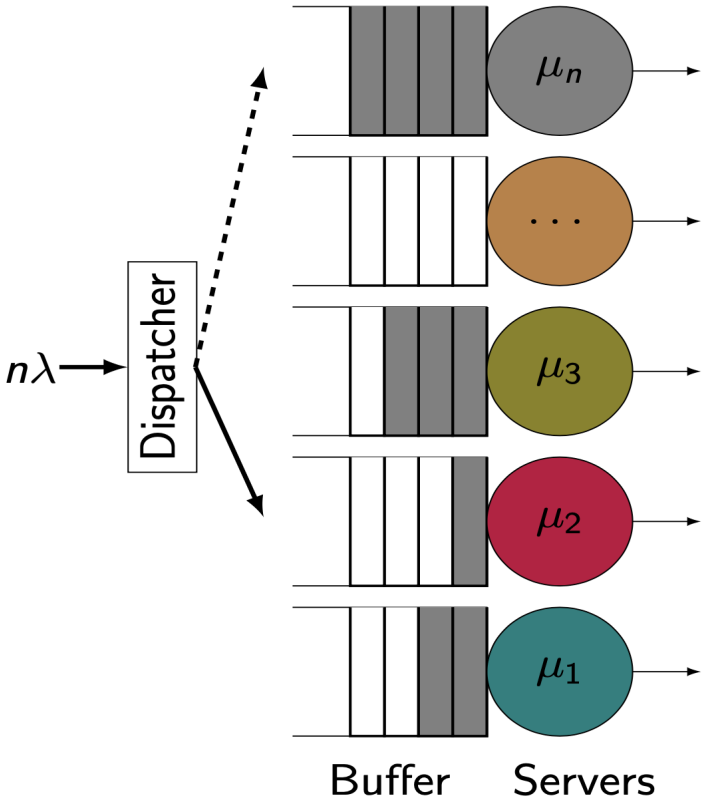
$$\lim_{n\to\infty} M^{(n)}(t) = m(t) \qquad\qquad (a.s.\ or\ in\ proba.)$$

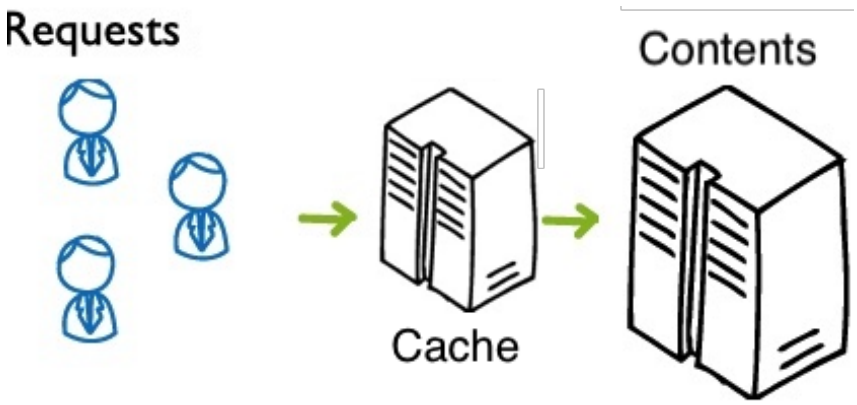
Yet, what often matters is the bias $\mathbb{E}\left[M^{(n)}(t)\right] - m(t)$.

- Bias $= \frac{1}{n}V(t) + O(\frac{1}{n^2})$ for smooth and homogeneous system.
- $V(t)$ can be computed in $O(S^3)$ `https://pypi.org/project/rmftool/`

Outline

Motivation for Heterogeneous Models





Zoom on a cache-replacement policy: RAND(c)

The cache i can contain c_i items.
Requests for item k arrive at rate λ_k.

- Requests trigger promotion.

Zoom on a cache-replacement policy: RAND(c)

The cache i can contain c_i items.
Requests for item k arrive at rate λ_k.

- Requests trigger promotion.

This is a pairwise interaction model:

- State of an object is its list.
- Transitions for objects k, k':

$$(i, i+1) \mapsto (i+1, i) \qquad \text{at rate } \lambda_k / c_{i+1}.$$

Mean field heterogeneous model

$M_s^{(n)}(t) = \frac{1}{n}\{\#\text{ objects in state } s \text{ at } t\}$ $\Rightarrow M^{(n)}$ is not Markovian.

Mean field heterogeneous model

~~$M_s^{(n)}(t) = \frac{1}{n}\{\#$ objects in state s at $t\}$ $\Rightarrow M^{(n)}$ is not Markovian.~~

Solution: represent model using indicators:

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

$\mathbf{X}^{(n)}$ is Markovian. Example, for the cache:

$$\mathbf{X}^{(n)} \to \mathbf{X}^{(n)} + e_{k,i+1} + e_{k',i} - e_{k,i} - e_{k',i+1} \text{ at rate } \frac{\lambda_k}{c_{i+1}} X_{k,i} X_{k',i+1}.$$

Heterogeneous mean field approximation

Similarly to the homogeneous case, we define the drift for X:

$$f^{(n)}(\mathbf{X}^{(n)}(t)) \approx \sum_{\substack{\text{set of jumps} \\ \text{from } \mathbf{X}^{(n)}(t)}} \text{jump} \times \text{jump rate}$$

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

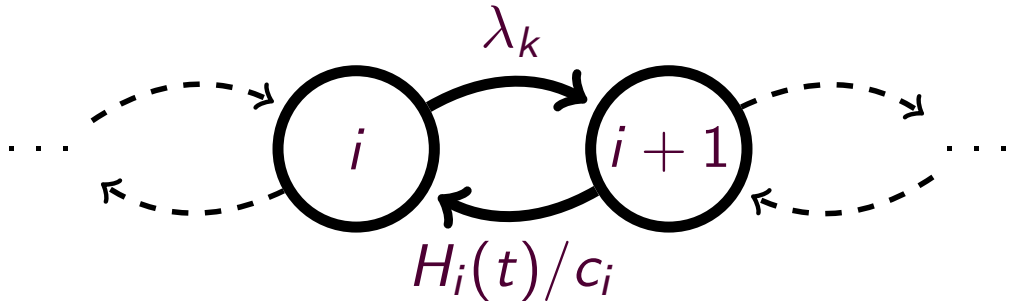
Heterogeneous mean field approximation

Similarly to the homogeneous case, we define the drift for X:

$$f^{(n)}(\mathbf{X}^{(n)}(t)) \approx \sum_{\substack{\text{set of jumps} \\ \text{from } \mathbf{X}^{(n)}(t)}} \text{jump} \times \text{jump rate}$$

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

Example (cache): Let $H_i(t) = \sum_k \lambda_k X_{k,i}(t)$ be the popularity in list i.



m-f approximation = use $\mathbb{E}[H_i(t)]$ to break dependencies.

Accuracy of the heterogeneous mean field

Theorem (Allmeier, G. 2022)

Assume that the rates r (from slide 8) are bounded, then there exists a computable constant $V^{(n)}(t) = O(1/n)$ such that:

$$\mathbb{E}\left[\mathbf{X}^{(n)}(t)\right] = x(t) + V(t) + O(1/n^2).$$

For the cache: this holds if $\lambda_k = O(1)$ and $c_i = \Theta(n)$.

What is the x and the V term and why this works?

Transitions are of the form:

$$\mathbf{X}^{(n)} \to \mathbf{X}^{(n)} + e_{k,j} + e_{k',j'} - e_{k,i} - e_{k',i'} \text{ at rate } \frac{1}{n} r_{ij,i'j'}^{(k,k')} X_{k,i} X_{k',i+1} \ ,$$

which can be written

$$\mathbf{X}^{(n)} \to \mathbf{X}^{(n)} + \ell \text{ at rate } \beta_\ell(x),$$

with $\beta_\ell(x) = \frac{1}{n} r_{ij,i'j'}^{(k,k')} X_{k,i} X_{k',i+1}$.

What is the x and the V term and why this works?

Transitions are of the form:

$$\mathbf{X}^{(n)} \to \mathbf{X}^{(n)} + e_{k,j} + e_{k',j'} - e_{k,i} - e_{k',i'} \text{ at rate } \frac{1}{n} r_{ij,i'j'}^{(k,k')} X_{k,i} X_{k',i+1} \ ,$$

which can be written

$$\mathbf{X}^{(n)} \to \mathbf{X}^{(n)} + \ell \text{ at rate } \beta_\ell(x),$$

with $\beta_\ell(x) = \frac{1}{n} r_{ij,i'j'}^{(k,k')} X_{k,i} X_{k',i+1}$.

x and V are defined as for the homogeneous case.

- Theorem: "it works!"

The proof holds by using a similar methodology as for the homogeneous case and carefully examining the reminder terms.

Illustration: Approximation Error for the cache model

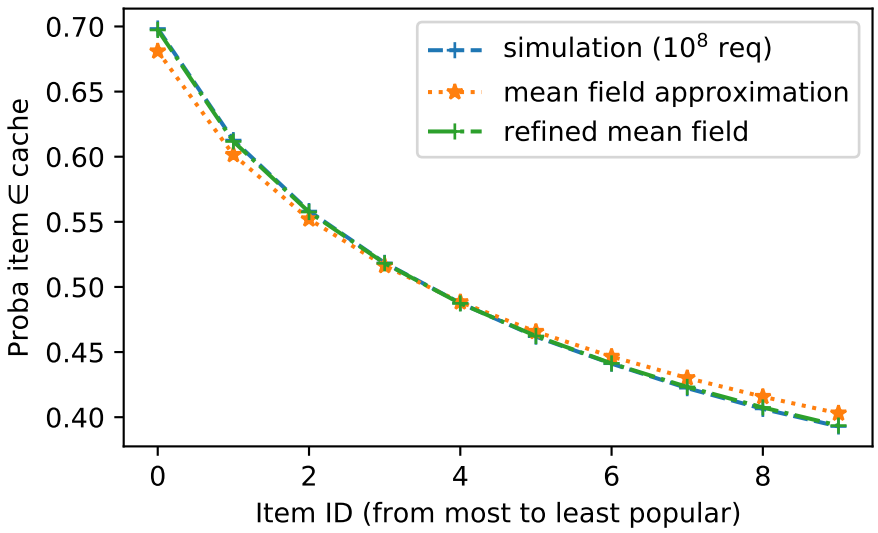
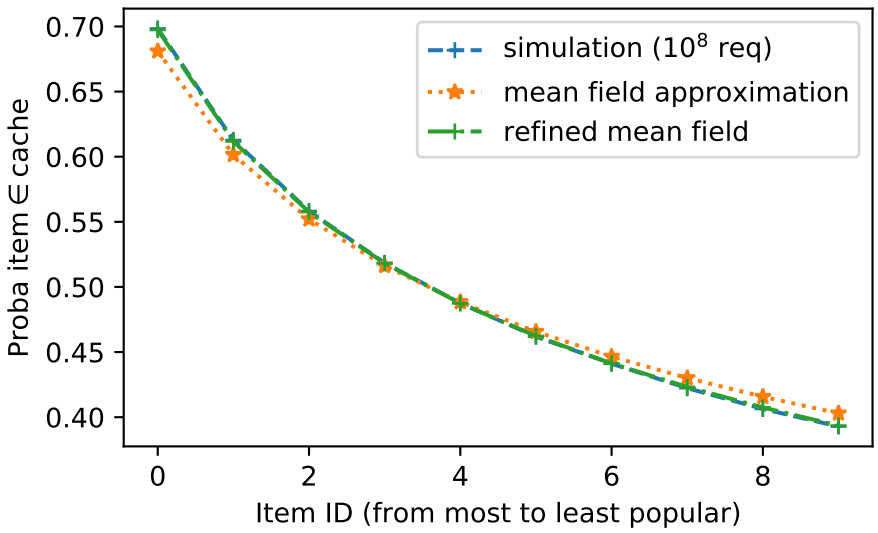


Illustration: Approximation Error for the cache model



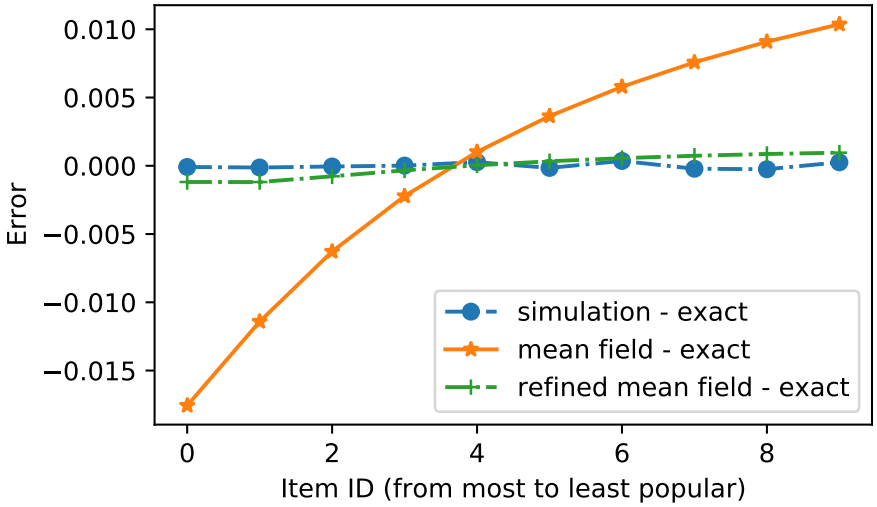
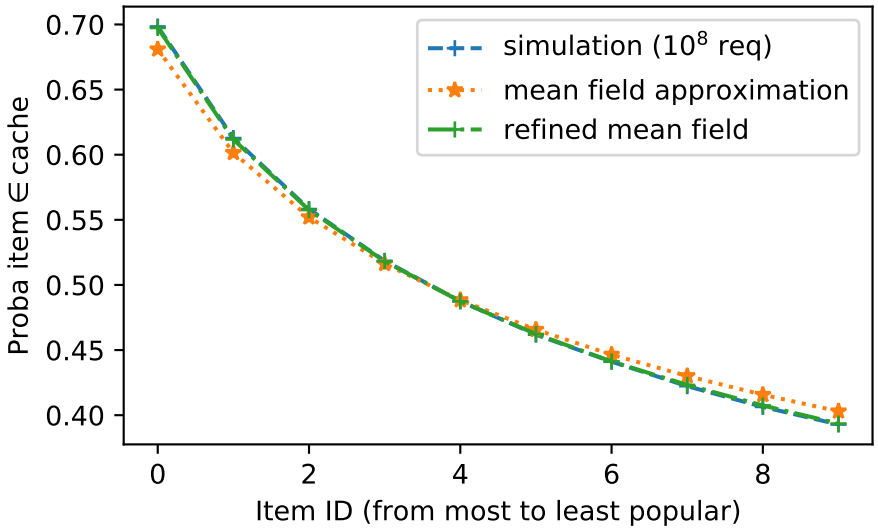
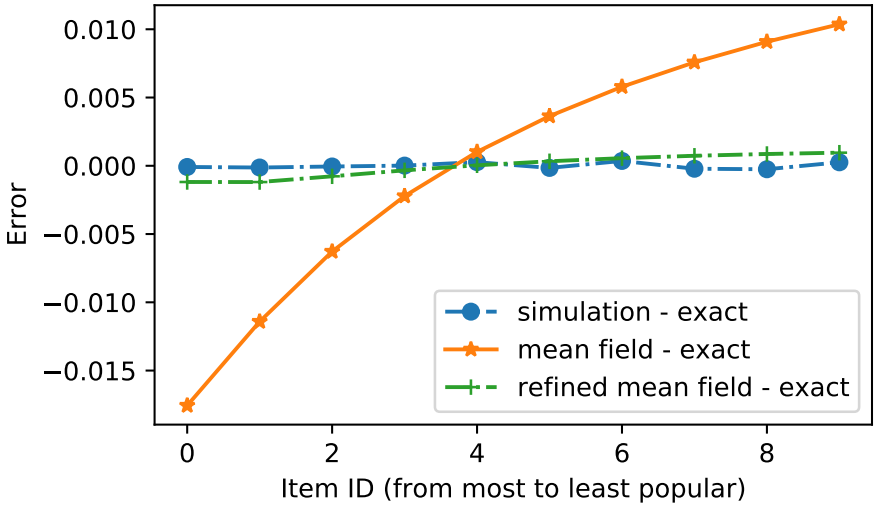


Illustration: Approximation Error for the cache model





	Mean field		Refined mean field		Simulation	
n	Error	(time)	Error	(time)	Error	(time)
10	0.0142	(10ms)	0.00197	(10ms)	0.00026	(4.3s)
30	0.0050	(14ms)	0.00022	(17ms)	0.00047	(4.9s)
50	0.0031	(17ms)	0.00008	(30ms)	0.00055	(5.7s)

For $n \geq 30$ refined mean field is more accurate than simulating 10^8 requests

Recap on heterogeneous models

Model: Object k, k' jump from states (i, i') to (j, j') at rate $\frac{1}{n} r_{ij,i'j'}^{(k,k')}$.

Our results: if the rates r are bounded:

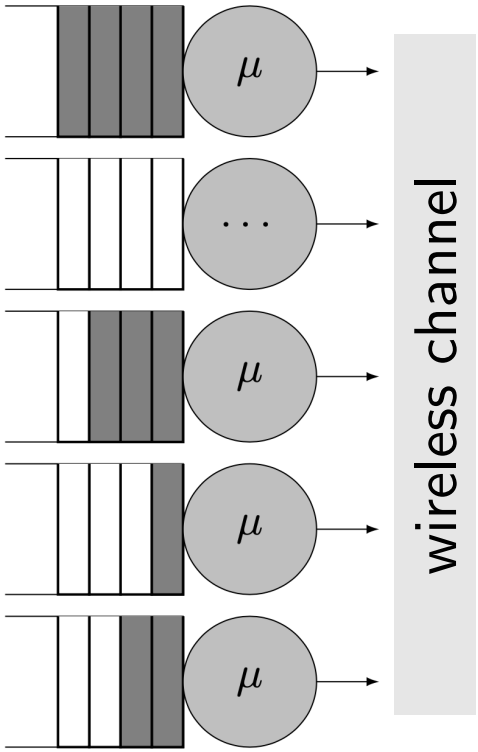
- Mean field approximation is $O(1/n)$-accurate.
- A refined approximation can be defined and is $O(1/n^2)$-accurate.

Outline

Multi-scale model

ex: Mean-field CSMA model

Cecchi, Borst, Leeuwaarden, Whiting 16



A server with a non-empty queue:

- Becomes active at rate μ if nobody is active.

Markovian model:

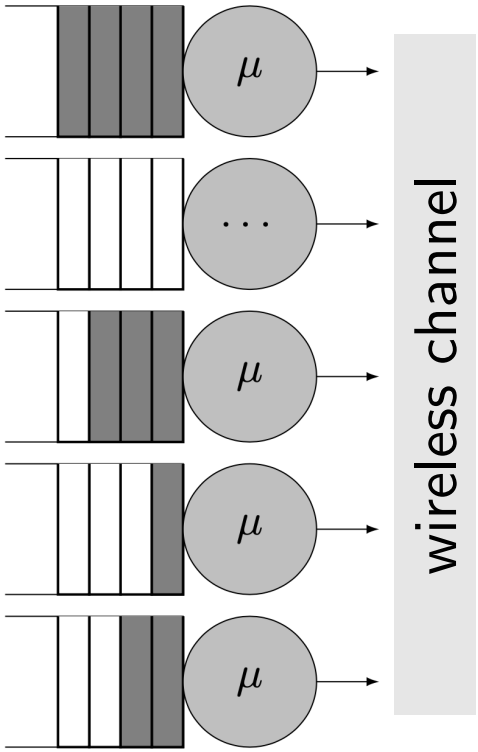
$$M^{(n)}(t), B(t),$$

where $B(t)$=channel's state.

Multi-scale model

ex: Mean-field CSMA model

Cecchi, Borst, Leeuwaarden, Whiting 16



A server with a non-empty queue:

- Becomes active at rate μ if nobody is active.

Markovian model:

$$M^{(n)}(t), B(t),$$

where $B(t)$=channel's state.

This is a two-scale model:

- $M^{(n)}(t)$ evolves by small jumps.
- $B^{(n)}(t)$ evolves fast.

Multi-scale model: mean field approximation

Classical approach:

- Drift $= f(m, b)$
- Mean field approximation: $\dot{m} = \sum_b f(m, b)\pi_b(m)$.

Multi-scale model: mean field approximation

Classical approach:

- Drift $= f(m, b)$
- Mean field approximation: $\dot{m} = \sum_b f(m, b)\pi_b(m)$.

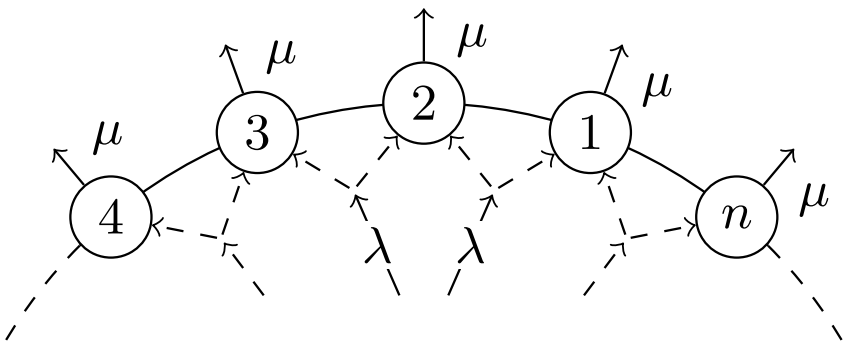
One can prove almost sure convergence by using: Cecchi et al 16, Ball at al 05

$$f(M^{(n)}(t), B^{(n)}(t)) \approx \sum_b f(M^{(n)}(t), b)\pi_b(M^{(n)}(t)).$$

Current work / open problem: for smooth systems

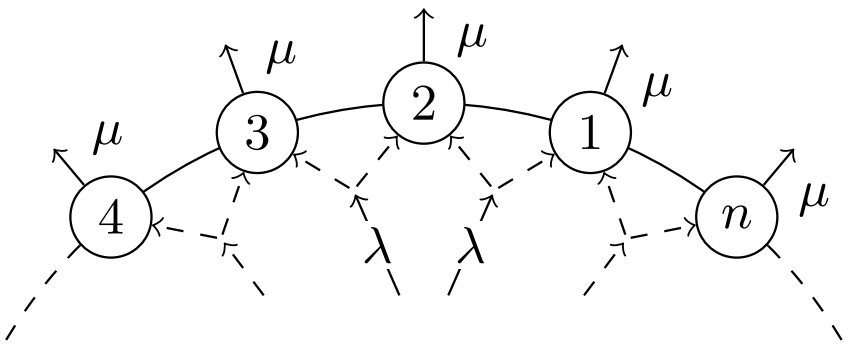
- The bias is $O(1/n)$
- Can we compute it?

Geometric aspects

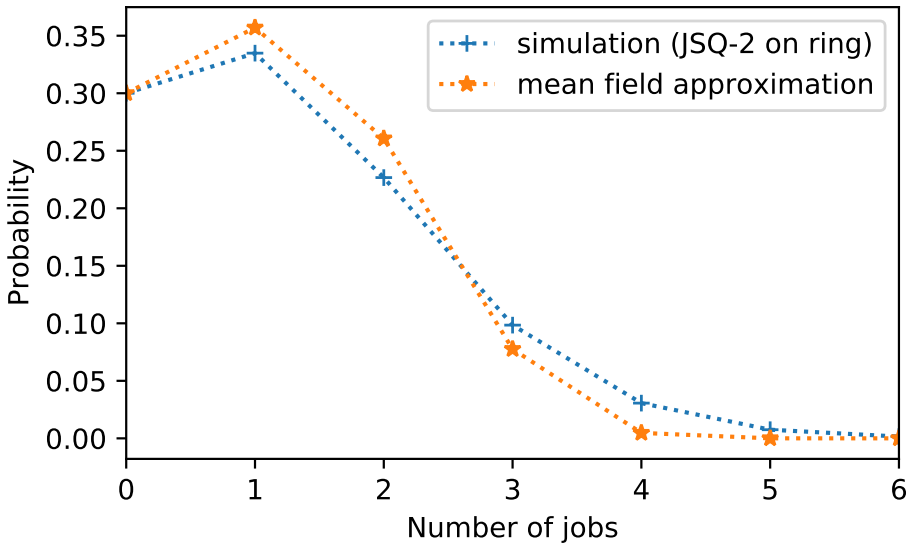


Example: Two-choice on a ring.

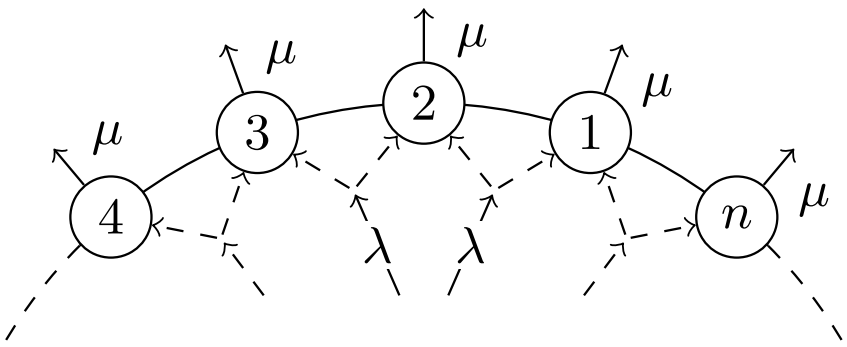
Geometric aspects



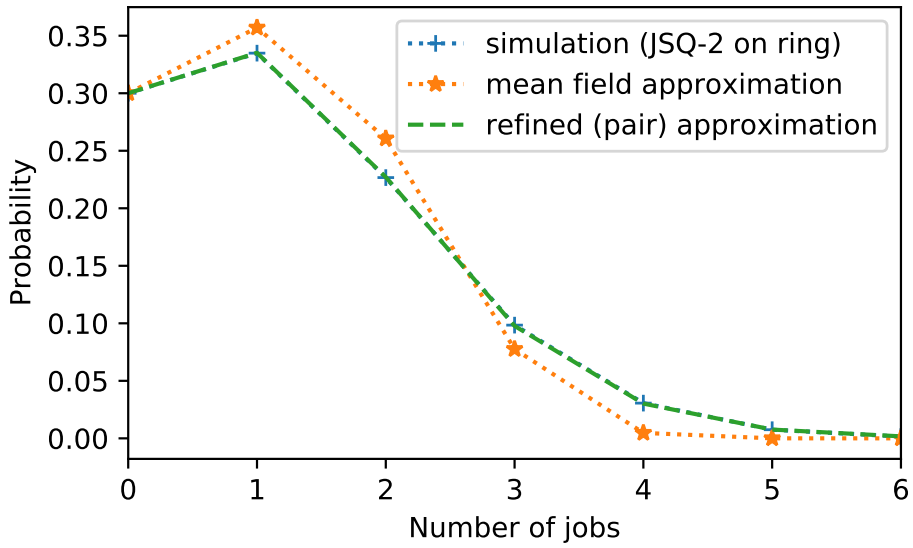
Example: Two-choice on a ring.



Geometric aspects



Example: Two-choice on a ring.



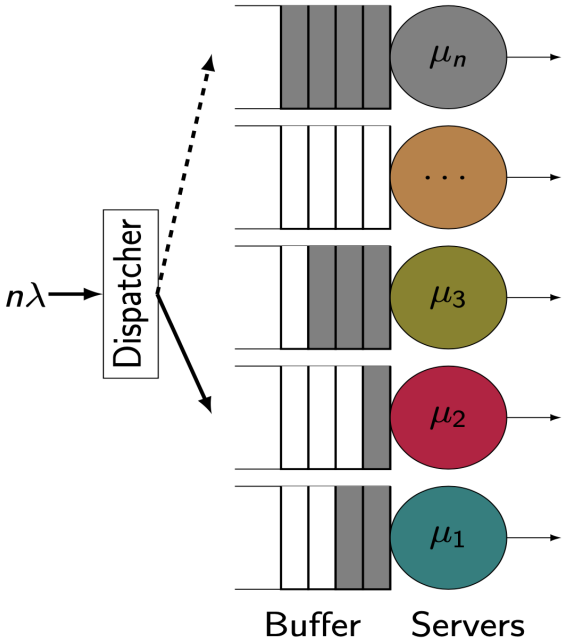
Pair-approximation is a moment closure technique: let x_i, y_{ij} and z_{ijk} be the proportions of node/edges/triplets with $i/(ij)/(ijk)$ jobs and assume:

$$z_{ijk} \approx y_{ji} y_{jk} / x_j .$$

Open questions:

- Why does this work?
- Related to replica mean field?

Non-markovian dynamics



Bramson, Lu, Prabhakar

Non-exponential service times.

- Asymptotically insensitive under PS.
- A mf-approx can be defined.

Open questions:

- Is the bias of order $O(1/n)$?
- How to use the covariance here?

Outline

Conclusion

Mean field approximation is a widely used heuristic.

- It consists in assuming independence.

We question its validity / accuracy.

Conclusion

Mean field approximation is a widely used heuristic.

- It consists in assuming independence.

We question its validity / accuracy.

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

Conclusion

Mean field approximation is a widely used heuristic.

- It consists in assuming independence.

We question its validity / accuracy.

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

Many open questions: geometric models, non-Markovian, controlled systems

More 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.
- Expected Values Estimated via Mean Field Approximation are O(1/N)-accurate by Gast. SIGMETRICS 2017.
- Transient and steady-state regime of a family of list-based cache replacement algorithms Gast and Van Houdt. SIGMETRICS 2015

Paper cited as open problems:

- Pair-approximation: The Power of Two Choices on Graphs: the Pair-Approximation is Accurate by Gast. Mama 2015.
- Replica mean field: Replica-mean-field limits for intensity-based neural networks by Baccelli and Taillefumier. 2019; The Pair-Replica-Mean-Field Limit for Intensity-based Neural Networks by Baccelli and Taillefumier. 2020.
- Two-scale: CSMA networks in a many-sources regime: A mean-field approach. Cecchi, Borst, van Leeuwaarden, Whiting. Infocom 2016
- 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