Approximating the bias of stochastic processes

With applications to stochastic approximation and mean-field limits.

Nicolas Gast (Inria, Grenoble), joint work with Sebastian Allmeier

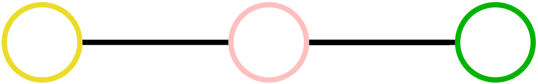
Inria Sophia, May 2024

Motivating example: Independent sets with arrivals

CSMA model from Cecchi et al. 2015

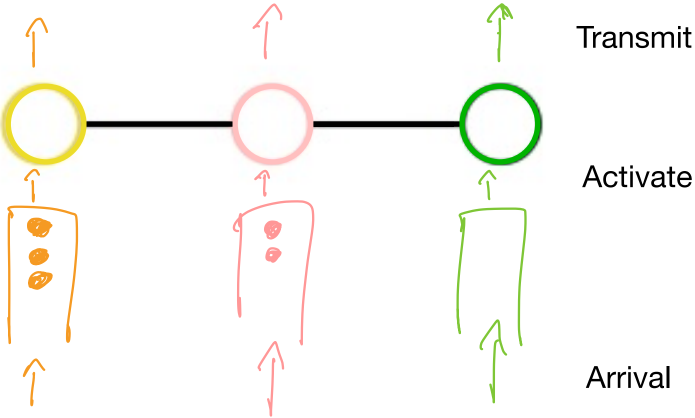
Motivating example: Independent sets with arrivals

CSMA model from Cecchi et al. 2015



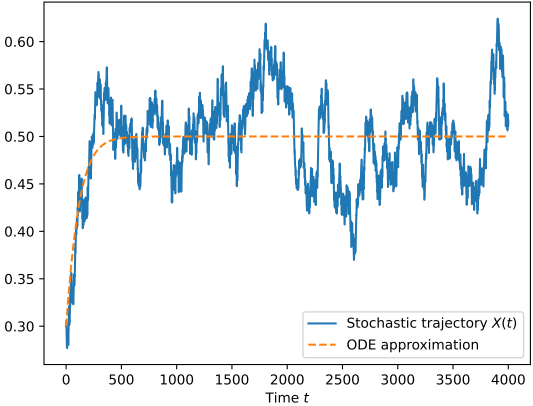
Motivating example: Independent sets with arrivals

CSMA model from Cecchi et al. 2015



Objective: estimate $\mathbf{P}[S_k = i]$.

The ODE method

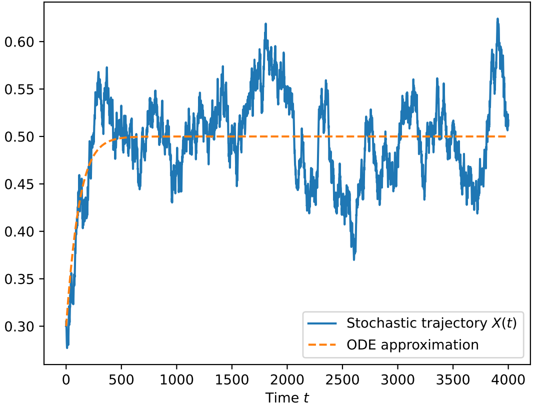


↙

1. Study a complex (e.g. queueing) system

(Mean field methods)

The ODE method . . . has (at least) two applications.



↙ ↘

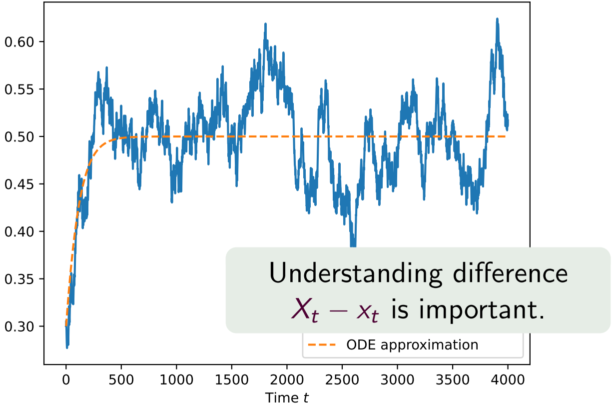
1. Study a complex (e.g. queueing) system

(Mean field methods)

2. Design an algorithm that solves $\bar{f}(x) = 0$.

(Stochastic approximation)

The ODE method . . . has (at least) two applications.



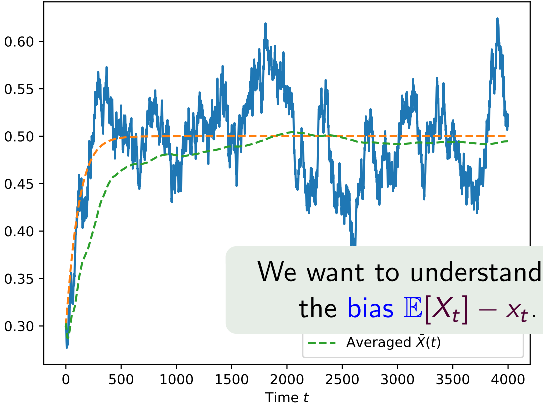
1. Study a complex (e.g. queueing) system

(Mean field methods)

2. Design an algorithm that solves $\bar{f}(x) = 0$.

(Stochastic approximation)

The ODE method ...has (at least) two applications.



We want to understanding the bias $\mathbb{E}[X_t] - x_t$.

1. Study a complex (e.g. queueing) system

(Mean field methods)

2. Design an algorithm that solves $\bar{f}(x) = 0$.

(Stochastic approximation)

Why is the bias important? (1/2)

... for mean field methods

System with N objects

$$X_i(t) = \frac{1}{N}\#\{\text{Objects in state } i \text{ at time } t\}$$

In steady-state:

$$\mathbf{P}\left[\text{An object is in state } i\right] = \mathbb{E}\left[X_i\right].$$

Why is the bias important? (1/2)

... for mean field methods

System with N objects

$$X_i(t) = \frac{1}{N}\#\{\text{Objects in state } i \text{ at time } t\}$$

In steady-state:

$$\mathbf{P}\left[\text{An object is in state } i\right] = \mathbb{E}\left[X_i\right].$$

Assume that we can construct some mean field approximation $\dot{x} = \bar{f}(x)$ with a fixed point x^*, then:

- How good does x_i^* approximate $\mathbf{P}\left[\text{An object is in state } i\right]$?

Why is the bias important? (2/2)

...for stochastic approximation is important

Model with constant step-size and Markovian noise:

$$\theta_{n+1} = \theta_n + \alpha \left(f(\theta_n, Y_n) + M_{n+1} \right)$$

where $Y_{n+1} \sim \mathbf{P}\left[\cdot \mid \theta_n, Y_n\right]$

Why is the bias important? (2/2)

... for stochastic approximation is important

Model with constant step-size and Markovian noise:

$$\theta_{n+1} = \theta_n + \alpha\left(f(\theta_n, Y_n) + M_{n+1}\right)$$

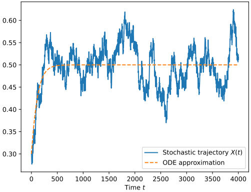
where $Y_{n+1} \sim \mathbf{P}\left[\cdot \mid \theta_n, Y_n\right]$, and define

$$\bar{\theta}_n = \frac{1}{n}\sum_{k=1}^{n} \theta_k.$$

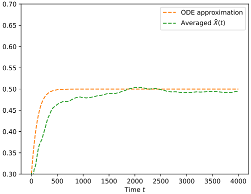
Assume that the ODE approximation: $\dot{\theta} = \bar{f}(\theta)$ has an attractor θ^*.

- How far is $\bar{\theta}_n$ from θ^* (for large n and small α?)

Reults in a nutshell

1. In general: $|X(t) - x(t)| = O(\sqrt{\frac{1}{N}})$ 

.

2. If $\bar{f}$ is *smooth*, then: $|\bar{X}(t) - x(t)| = O(\frac{1}{N})$ 

.

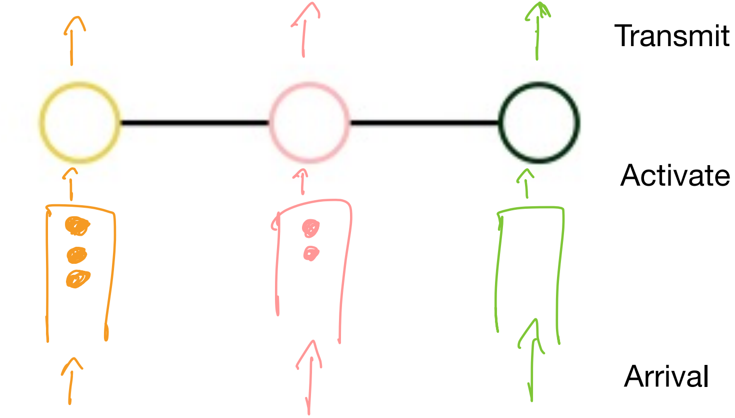
3. Results for the stochastic approximation with constant step size $\alpha = 1/N$ corresponds to results for a mean field interacting model with N objects.

Outline

1. Mean field interaction models with a shared resource
2. Elements of Proof (Stein's method)
3. What about stochastic approximation?
4. Conclusion

Replica mean field model with a fast varying environment

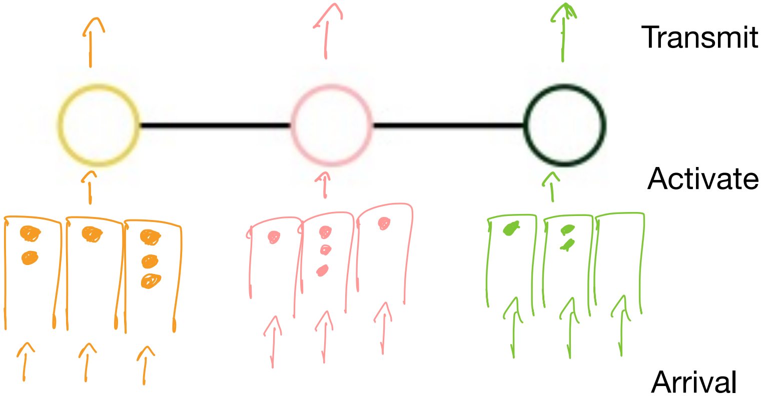
Independent sets with arrivals, CSMA model from Cecchi et al. 2015



Objective: estimate $\mathbf{P}[S_k = i]$

Replica mean field model with a fast varying environment

Independent sets with arrivals, CSMA model from Cecchi et al. 2015



Objective: estimate $\mathbf{P}[S_k = i]$

Scaling: *replica*

- N severs per node
- Transmission rate $\times N$.

Mean field model

Population of N objects.

- Object k has a state $S_k(t) \in \mathcal{S}$.

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

Model

- Object k jumps from i to j at rate $Q_{ij}(\mathbf{X})$

Mean field model . . . with a shared resource

Population of N objects.

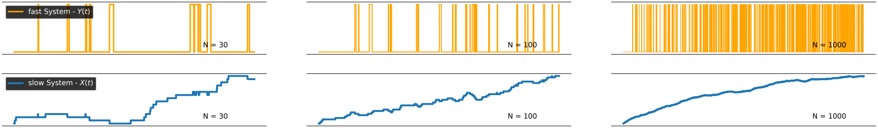
- Object k has a state $S_k(t) \in \mathcal{S}$.

$$X_i = \text{ fraction of objects in state } i.$$
$$Y = \text{“activation set''}$$

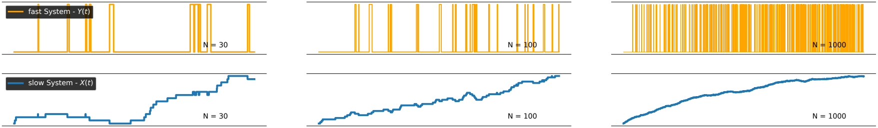
Model

- Object k jumps from i to j at rate $Q_{ij}(\mathbf{X}, \mathbf{Y})$
- Resource $\mathbf{Y}$ jumps from y to y' at rate $NK_{y,y'}(\mathbf{X})$.

We use two approximations to construct a fluid limit.



We use two approximations to construct a fluid limit.



"Average" mean field approximation

Let $\pi_y(\mathbf{x})$ be the stationary distribution of $K(\mathbf{x})$. We define:

$$\bar{Q}(\mathbf{x}) = \sum_y \pi_y(\mathbf{x}) Q(\mathbf{x}, y).$$

The mean field approximation is the solution of the ODE:

$$\dot{\mathbf{x}} = \mathbf{x}\bar{Q}(\mathbf{x}),$$

The approximation is asymptocally exact. Its bias is v/N.

Assume that

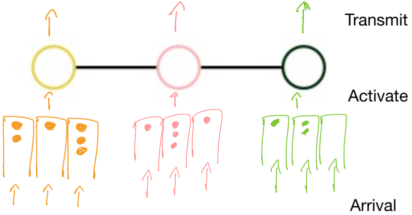
- K is unichain for all $\mathbf{x}$.
- K and Q are twice differenciable
- $\dot{\mathbf{x}} = \mathbf{x}\bar{Q}(\mathbf{x})$ has a unique attractor x^*.

Theorem

There exists a *computable* V such that, in steady-state:

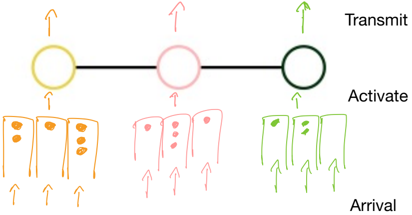
$$\mathbf{P}\left[S_k = i\right] = \underbrace{\underbrace{x_i^*}_{\text{mean field approximation}} + \frac{1}{N} V_i}_{\text{refined approximation}} + O(1/N^2).$$

Illustration of the theorem

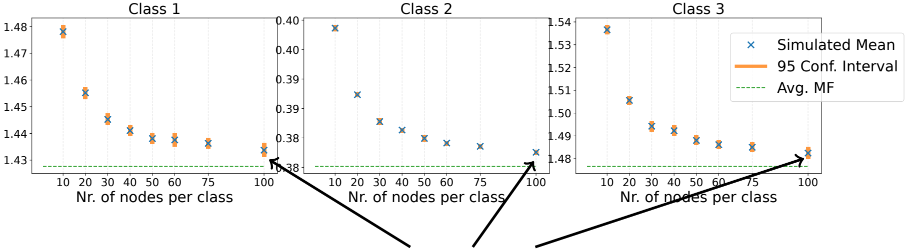


Transmission rates	1.4	1.3	1.7
Activation rates	1.2	2	1.5
Arrival rates	0.5	0.2	0.5

Illustration of the theorem

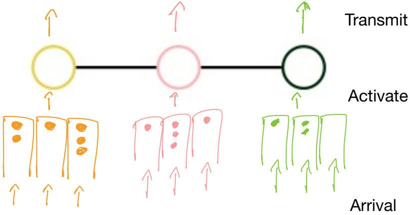


Transmission rates	1.4	1.3	1.7
Activation rates	1.2	2	1.5
Arrival rates	0.5	0.2	0.5

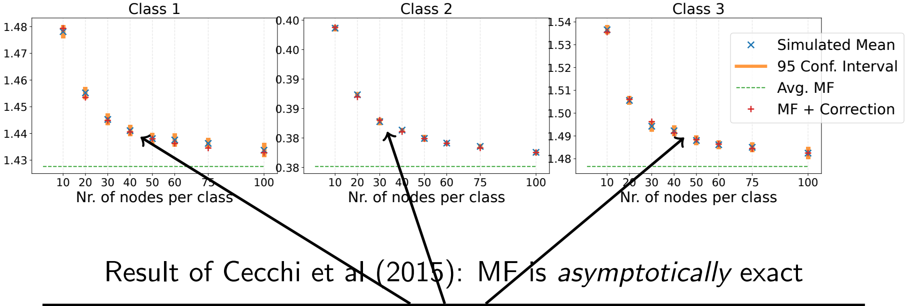


Result of Cecchi et al (2015): MF is *asymptotically* exact

Illustration of the theorem



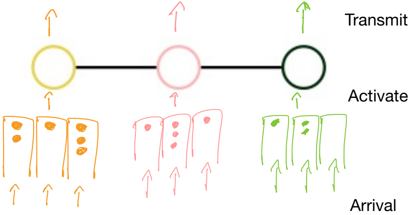
Transmission rates	1.4	1.3	1.7
Activation rates	1.2	2	1.5
Arrival rates	0.5	0.2	0.5



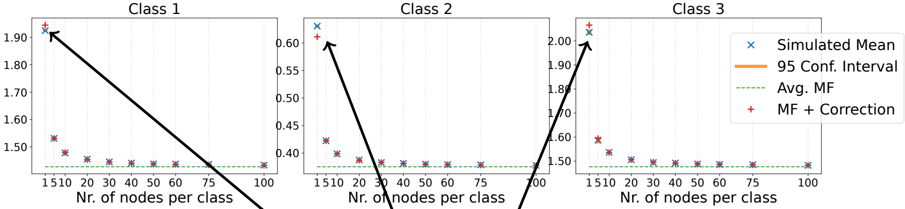
Result of Cecchi et al (2015): MF is *asymptotically* exact

Our results: Accuracy is $O(1/N)$. MF+correction is almost exact

Illustration of the theorem



Transmission rates	1.4	1.3	1.7
Activation rates	1.2	2	1.5
Arrival rates	0.5	0.2	0.5



Result of Cecchi et al (2015): MF is *asymptotically* exact

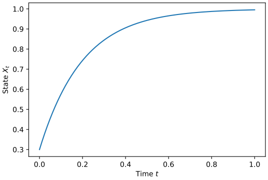
Our results: Accuracy is $O(1/N)$. MF+correction is almost exact

Even for $N = 1$.

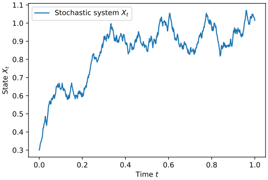
Outline

To compare X_t and ODE, we study infinitesimal changes

The generator approach



ODE $\phi_t(X_0)$



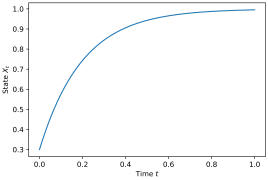
X_t

We want to compare:

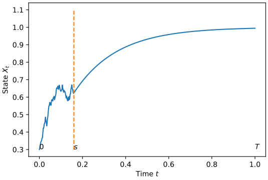
$$\mathbb{E}[X_t] - \phi_t(X_0)$$

To compare X_t and ODE, we study infinitesimal changes

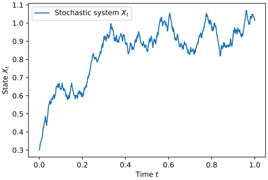
The generator approach



ODE $\phi_t(X_0)$



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



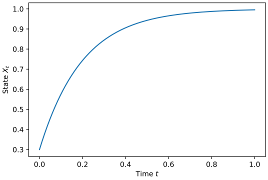
X_t

We want to compare:

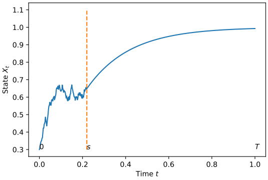
$$\mathbb{E}[X_t] - \phi_t(X_0)$$

To compare X_t and ODE, we study infinitesimal changes

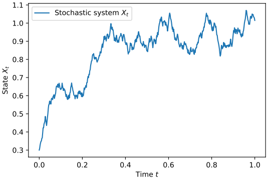
The generator approach



ODE $\phi_t(X_0)$



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



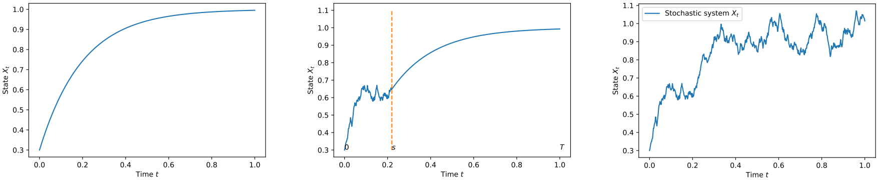
X_t

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 compare X_t and ODE, we study infinitesimal changes

The generator approach



ODE $\phi_t(X_0)$ ODE $\phi_{t-s}(X_s)$ X_t

We want to compare:

$$\mathbb{E}\left[X_\infty\right] - \phi_\infty(X_0) = \int_0^\infty \mathbb{E}\left[\frac{d}{ds}\phi_{t-s}(X_s)\right] ds$$
$$= \int_0^\infty (G^{\mathrm{sto}} - G^{\mathrm{ODE}})\phi_{t-s}(X_s) ds.$$

If Y is not here, we can directly use Stein's method

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

1. $$G^{\mathrm{sto}} h(X) = \sum_{i,j} (h(X + \frac{1}{N}(e_j - e_i)) - h(X)) N x_i Q_{ij}(x)$$

If Y is not here, we can directly use Stein's method

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

1. $$G^{\mathrm{sto}} h(X) = \sum_{i,j} (h(X + \frac{1}{N}(e_j - e_i)) - h(X)) N x_i Q_{ij}(x)$$
$$= \underbrace{\nabla h \cdot x Q(X)}_{\text{Generator of ODE } \dot{x} = xQ(x).} + \underbrace{O(1/N)}_{\text{if } h \text{ is } C^1}.$$

If Y is not here, we can directly use Stein's method

and introduce a Poisson equation for the slow system

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

1. $$G^{\mathrm{sto}}h(X) = \sum_{i,j}(h(X + \frac{1}{N}(e_j - e_i)) - h(X))Nx_i Q_{ij}(x)$$
$$= \underbrace{\nabla h \cdot xQ(X)}_{\text{Generator of ODE } \dot{x} = xQ(x).} + \underbrace{O(1/N)}_{\text{if } h \text{ is } C^1}.$$

2. $\mathbb{E}\left[G^{\mathrm{sto}}h(X)\right] = 0$ if X is in steady-state.

If Y is not here, we can directly use Stein's method

and introduce a Poisson equation for the slow system

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

1. $$G^{\mathrm{sto}} h(X) = \sum_{i,j} (h(X + \frac{1}{N}(e_j - e_i)) - h(X)) N x_i Q_{ij}(x)$$
$$= \underbrace{\nabla h \cdot xQ(X)}_{\text{Generator of ODE } \dot{x} = xQ(x).} + \underbrace{O(1/N)}_{\text{if } h \text{ is } C^1}.$$

2. $\mathbb{E}\left[G^{\mathrm{sto}} h(X)\right] = 0$ if X is in steady-state.

Let G be such that $\nabla G \cdot xQ(x) = x - x^*$ (Poisson equation). We have:

$$\begin{aligned}
\mathbb{E}\left[X - x^*\right] &= \mathbb{E}\left[\nabla G \cdot XQ(X)\right] \\
&= \mathbb{E}\left[\nabla G \cdot XQ(X) - G^{\mathrm{sto}} G\right] && \text{(by (2))} \\
&= O(1/N) && \text{(by (1))}.
\end{aligned}$$

When Y is here, we need to treat the fast system

Let $h : \mathcal{X} \to \mathbb{R}$ be a test function. We have:

$$\begin{aligned} G^{\mathrm{sto}} h(X,Y) &= \sum_{i,j} (h(X + \frac{1}{N}(e_j - e_i)) - h(X)) N x_i Q_{ij}(X,Y) \\ &= \underbrace{\nabla h \cdot XQ(X,Y)}_{\neq G^{\mathrm{ODE}} = \nabla h \cdot \bar{Q}(x).} + O(1/N), \end{aligned}$$

When Y is here, we need to treat the fast system

Let $h : \mathcal{X} \to \mathbb{R}$ be a test function. We have:

$$G^{\mathrm{sto}} h(X,Y) = \sum_{i,j} (h(X + \frac{1}{N}(e_j - e_i)) - h(X)) N x_i Q_{ij}(X,Y)$$
$$= \underbrace{\nabla h \cdot XQ(X,Y)}_{\neq G^{\mathrm{ODE}} = \nabla h \cdot \bar{Q}(x).} + O(1/N),$$

We are left with $Q(X,Y) - \bar{Q}(X)$.

When Y is here, we need to treat the fast system

and introduce a Poisson equation for the fast system.

Let $h : \mathcal{X} \to \mathbb{R}$ be a test function. We have:

$$G^{\mathrm{sto}}h(X,Y) = \sum_{i,j}(h(X + \frac{1}{N}(e_j - e_i)) - h(X))Nx_iQ_{ij}(X,Y)$$
$$= \underbrace{\nabla h \cdot XQ(X,Y)}_{\neq G^{\mathrm{ODE}} = \nabla h \cdot \bar{Q}(x).} + O(1/N),$$

We are left with $Q(X,Y) - \bar{Q}(X)$.

Lemma: There exists a K^+ that is C^2 such that for all $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$:

$$h(X,Y) - \bar{h}(X) = K(x)K^+(x)h(X,Y).$$

Rapping up the proof

Let $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a test function.

$$G^{\mathrm{sto}} h(X,Y) = \underbrace{NK(X)h(X,Y)}_{\text{fast}} + \underbrace{\nabla_x h \cdot XQ(X,Y)}_{\text{slow}} + O(1/N)$$

Hence, $K^{fast} - \frac{1}{N} G^{\mathrm{sto}} = o(1/N)$ if h is C^1.

Rapping up the proof

Let $h : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be a test function.

$$G^{\text{sto}} h(X, Y) = \underbrace{NK(X)h(X, Y)}_{\text{fast}} + \underbrace{\nabla_x h \cdot XQ(X, Y)}_{\text{slow}} + O(1/N)$$

Hence, $K^{fast} - \frac{1}{N} G^{\text{sto}} = o(1/N)$ if h is C^1.

This shows that in steady-state:

$$\begin{aligned}
\mathbb{E}\left[h(X) - \bar{h}(X)\right] &= \mathbb{E}\left[K(x)K^{+}(x)h(X, Y)\right] \\
&= \mathbb{E}\left[(K(x) - \frac{1}{n} G^{\text{sto}})K^{+}(x)h(X, Y)\right] && \text{(steady-state).} \\
&= O(1/N) && \text{(expansion above).}
\end{aligned}$$

Outline

Stochastic approximation with Markovian noise

... is similar to our mean field model with a fast-varying environment

Recurrence of the form:

$$\theta_{n+1} = \theta_n + \alpha\left(f(\theta_n, Y_n) + M_{n+1}\right)$$

where $Y_{n+1} \sim \mathbf{P}\left[\cdot \mid \theta_n, Y_n\right]$.

Stochastic approximation with Markovian noise

... is similar to our mean field model with a fast-varying environment

Recurrence of the form:

$$\theta_{n+1} = \theta_n + \alpha \left(f(\theta_n, Y_n) + M_{n+1} \right)$$

where $Y_{n+1} \sim \mathbf{P}\left[\cdot \mid \theta_n, Y_n\right]$.

For our mean field model with a resource:

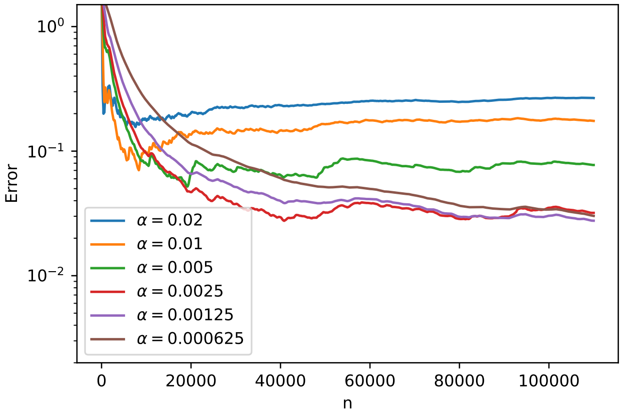
$$X(t+\frac{1}{N}) = X(t) + \frac{1}{N}(xQ(x,Y) + \underbrace{N(X(t+\frac{1}{N}) - X(t)) - xQ(x,Y)}_{:=M(t+\frac{1}{N}) \text{ and } \mathbb{E}\left[M(t+\frac{1}{N})\cdot|\mathcal{F}(t)\right]\approx 0}).$$

Hence, we can use similar proofs.

Stochastic approximation: results

Under "smoothness" conditions, there exists V such that

$$\limsup_{n\to\infty} \bar{\theta}_n = \theta^* + V\alpha + O(\alpha^2).$$



Here: Error $= \bar{\theta}_n - \theta^*$

We can extrapolate V by using two step-sizes α and 2α

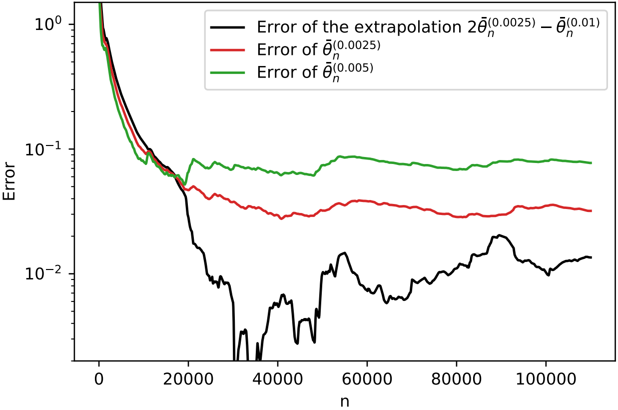
$$\bar{\theta}_n^{(\alpha)} = \theta^* + V\alpha + O(\alpha^2)$$
$$\bar{\theta}_n^{(2\alpha)} = \theta^* + V2\alpha + O(\alpha^2)$$

We can extrapolate V by using two step-sizes α and 2α

$$\bar{\theta}_n^{(\alpha)} = \theta^* + V\alpha + O(\alpha^2)$$
$$\bar{\theta}_n^{(2\alpha)} = \theta^* + V2\alpha + O(\alpha^2)$$

Hence:

$$2\bar{\theta}_n^{(\alpha)} - \bar{\theta}_n^{(2\alpha)} = \theta^* + O(\alpha^2).$$



Outline

Conclusion

We study the accuracy of mean field approximation for two time-scale.

- The bias is of order $O(1/N)$. It can be computed.
- This also works for most "smooths" models (e.g., heterogeneous).

Conclusion

We study the accuracy of mean field approximation for two time-scale.

- The bias is of order $O(1/N)$. It can be computed.
- This also works for most "smooths" models (e.g., heterogeneous).

Two-timescale models:

- Shared resource or synchronization (e.g., CSMA)
- Q-learning type algorithm: Stochastic approximation algorithms with Markovian noise. Huo et al. 2023

Conclusion

We study the accuracy of mean field approximation for two time-scale.

- The bias is of order $O(1/N)$. It can be computed.
- This also works for most "smooths" models (e.g., heterogeneous).

Two-timescale models:

- Shared resource or synchronization (e.g., CSMA)
- Q-learning type algorithm: Stochastic approximation algorithms with Markovian noise. Huo et al. 2023

Many open questions: non-smooth, (sparse) geometric models, non-Markovian.

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

References

Results on which this talk is based:

- Bias and Refinement of Multiscale Mean Field Models. Allmeier, Gast, 2022. Sigmetrics 2023.
- CSMA networks in a many-sources regime: A mean-field approach. Cecchi, Borst, van Leeuwaarden, Whiting. Infocom 2016.
- Results on stochastic approximation: preprint (email me if you want the preprint)

Q-learning and bias:

- Bias and Extrapolation in Markovian Linear Stochastic Approximation with Constant Stepsizes. Dongyan Huo, Yudong Chen, Qiaomin Xie. Sigmetrics 2023.

Related refined mean-field approximation papers:

- 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.