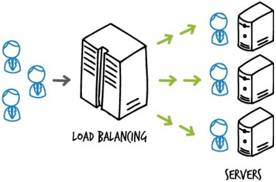
Bias of fluid / mean field approximation

Poisson equation, averaging methods and two-timescale processes

Nicolas Gast (Inria, Univ. Grenoble Alpes)

February 2025

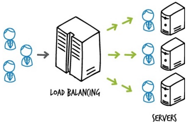
Motivation: Studying interacting particle systems



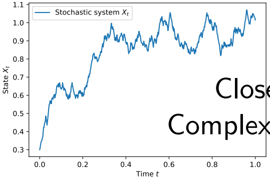
- Stochastic models are complex.

Fluid / mean field approximation simplifies the analysis

n servers



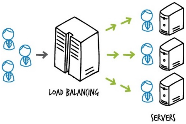
stochastic
model



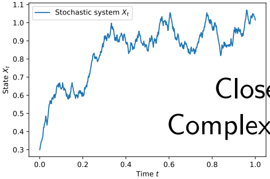
Close form?
Complex to analyze

Fluid / mean field approximation simplifies the analysis

n servers



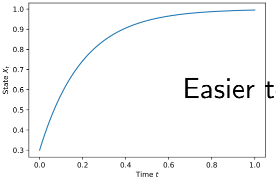
stochastic model



Close form?

Complex to analyze

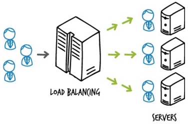
fluid



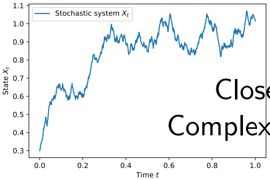
Easier to analyze

Fluid / mean field approximation simplifies the analysis

n servers



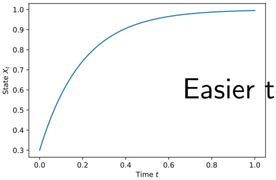
stochastic model



Close form?
Complex to analyze

Question: What is the error made?

fluid



Easier to analyze

Objective of this lecture

Provide tools to analyze the error of fluid approximations.

1. Introduction to stochastic generators (with examples)
2. Connecting generators and convergence of stochastic processes
3. Applications to density dependent population processes
4. Steady-state convergence and Stein's method

(Extensions: refined approximations, averaging methods.)

(Main) Related work

- Kurtz, 70s.
  - Fluid limits, diffusion limits (mostly transient regime)
- 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

Outline

1. Generators
   - Definitions
   - Examples
   - Link between generators and semi-groups
2. Connecting generators and convergence of stochastic processes
3. Application to density dependent processes
4. Steady-state and Stein's method
5. Conclusion

The semi-group operator of a stochastic process

In what follows, $(X_t)_t$ will denote a continuous time Markov process taking values in $\mathcal{X} \subset \mathbb{R}^d$.

We denote by T_t the semi-group of the stochastic $(X_t)_t$. For a function $h : \mathcal{E} \to \mathbb{R}$, it associates a function $T_t h$ defined by:

$$T_t h(x) = \mathbb{E}\left[h(X_t) \mid X_0 = x\right]$$

The semi-group operator of a stochastic process

In what follows, $(X_t)_t$ will denote a continuous time Markov process taking values in $\mathcal{X} \subset \mathbb{R}^d$.

We denote by T_t the semi-group of the stochastic $(X_t)_t$. For a function $h : \mathcal{E} \to \mathbb{R}$, it associates a function $T_t h$ defined by:

$$T_t h(x) = \mathbb{E}\left[h(X_t) \mid X_0 = x\right]$$

Note that $(T_t)_t$ is a semi-group:

- $T_0 h = h$
- For $t, s \geq 0$: $T_{t+s} = T_t \circ T_s$.
- $\lim_{t \to 0} T_t h - h = 0$.

The semi-group operator of a stochastic process

In what follows, $(X_t)_t$ will denote a continuous time Markov process taking values in $\mathcal{X} \subset \mathbb{R}^d$.

We denote by T_t the semi-group of the stochastic $(X_t)_t$. For a function $h : \mathcal{E} \to \mathbb{R}$, it associates a function $T_t h$ defined by:

$$T_t h(x) = \mathbb{E}\left[h(X_t) \mid X_0 = x\right]$$

Note that $(T_t)_t$ is a semi-group:

- $T_0 h = h$
- For $t, s \geq 0$: $T_{t+s} = T_t \circ T_s$.
- $\lim_{t \to 0} T_t h - h = 0$.

Abuse of notation: $T_t(x) = T_t \mathrm{Id}(x) = \mathbb{E}\left[X_t \mid X_0 = x\right]$

Infinitesimal generators

A function h is said to be in the domain of the generator if the following limit exists:

$$Ah := \lim_{t \to 0} \frac{T_t h - h}{t}$$

Infinitesimal generators

A function h is said to be in the domain of the generator if the following limit exists:

$$Ah := \lim_{t \to 0} \frac{T_t h - h}{t}$$

The infinitesimal generator of the stochastic process, denoted by A, is:

$$A = \lim_{t \to 0} \frac{T_t - T_0}{t}.$$

Examples. 1: discrete-state processes

- A Poisson process of intensity λ: $X_t \in \mathbb{Z}^+$.

$$T_t h(x) =$$

$$Ah(x) =$$

Examples. 1: discrete-state processes

- A Poisson process of intensity λ: $X_t \in \mathbb{Z}^+$.

$$
\begin{aligned}
T_t h(x) &= \sum_{k=0}^{\infty} \mathbf{P}\left[X_t = x + k \mid X_0 = x\right] h(x+k) \\
&= \sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{-\lambda}}{k!} h(x+k) \\
Ah(x) &= \lambda(h(x+1) - h(x)).
\end{aligned}
$$

Examples. 1: discrete-state processes

- A Poisson process of intensity λ: $X_t \in \mathbb{Z}^+$.

$$
\begin{aligned}
T_t h(x) &= \sum_{k=0}^{\infty} \mathbf{P}\left[X_t = x + k \mid X_0 = x\right] h(x+k) \\
&= \sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{-\lambda}}{k!} h(x+k) \\
Ah(x) &= \lambda(h(x+1) - h(x)).
\end{aligned}
$$

- A continuous time Markov chain $(X_t)_t$ with state space $\mathcal{E}$ and kernel K.

Examples. 1: discrete-state processes

- A Poisson process of intensity λ: $X_t \in \mathbb{Z}^+$.

$$T_t h(x) = \sum_{k=0}^{\infty} \mathbf{P}\left[X_t = x + k \mid X_0 = x\right] h(x+k)$$
$$= \sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{-\lambda}}{k!} h(x+k)$$
$$Ah(x) = \lambda(h(x+1) - h(x)).$$

- A continuous time Markov chain $(X_t)_t$ with state space $\mathcal{E}$ and kernel K.

$$Ah(x) = \sum_{y \in \mathcal{X}} (h(y) - h(x)) K_{xy}.$$

Examples. 2: continuous-state processes

- A differential equation: $\dot{x} = f(x)$.

Examples. 2: continuous-state processes

- A differential equation: $\dot{x} = f(x)$.

$$
\begin{aligned}
T_t(x) & \text{ is the solution of the ODE at time } t \text{ such that } X_0 = x \\
T_t h(x) &= f(T_t(x)) \\
Ah(x) &= Dh(x) f(x).
\end{aligned}
$$

Examples. 2: continuous-state processes

- A differential equation: $\dot{x} = f(x)$.

$$\begin{aligned} T_t(x) &\text{ is the solution of the ODE at time } t \text{ such that } X_0 = x \\ T_t h(x) &= f(T_t(x)) \\ Ah(x) &= Dh(x)f(x). \end{aligned}$$

- A Brownian motion (in $\mathbb{R}$) with variance σ^2

Examples. 2: continuous-state processes

- A differential equation: $\dot{x} = f(x)$.

$$\begin{aligned} T_t(x) &\text{ is the solution of the ODE at time } t \text{ such that } X_0 = x \\ T_t h(x) &= f(T_t(x)) \\ Ah(x) &= Dh(x)f(x). \end{aligned}$$

- A Brownian motion (in $\mathbb{R}$) with variance σ^2

$$Ah(x) = \frac{1}{2}\sigma^2 h''(x).$$

Examples. 2: continuous-state processes

- A differential equation: $\dot{x} = f(x)$.

$$\begin{aligned} & T_t(x) \text{ is the solution of the ODE at time } t \text{ such that } X_0 = x \\ T_t h(x) &= f(T_t(x)) \\ Ah(x) &= Dh(x)f(x). \end{aligned}$$

- A Brownian motion (in $\mathbb{R}$) with variance σ^2

$$Ah(x) = \frac{1}{2}\sigma^2 h''(x).$$

- A The Ornstein–Uhlenbeck process on $\mathbb{R}$ which satisfies the stochastic differential equation $\theta(\mu - X_t)dt + \sigma dB_t$:

Examples. 2: continuous-state processes

- A differential equation: $\dot{x} = f(x)$.

$$\begin{aligned} T_t(x) &\text{ is the solution of the ODE at time } t \text{ such that } X_0 = x \\ T_t h(x) &= f(T_t(x)) \\ Ah(x) &= Dh(x)f(x). \end{aligned}$$

- A Brownian motion (in $\mathbb{R}$) with variance σ^2

$$Ah(x) = \frac{1}{2}\sigma^2 h''(x).$$

- A The Ornstein–Uhlenbeck process on $\mathbb{R}$ which satisfies the stochastic differential equation $\theta(\mu - X_t)dt + \sigma dB_t$:

$$Ah(x) = \theta(\mu - x)h'(x) + \frac{\sigma^2}{2}h''(x).$$

Link between generators and semi-groups

Informally: $T_t = \exp(At)$.

Theorem

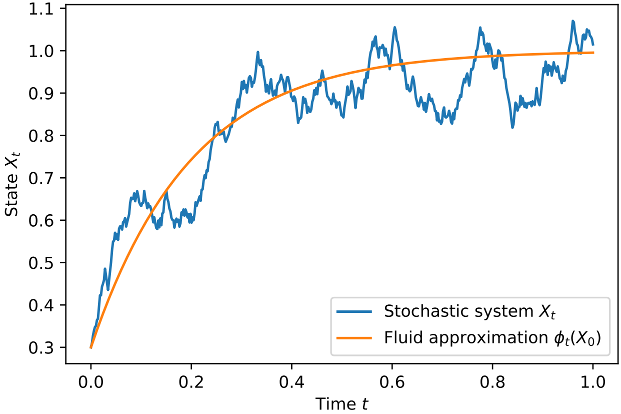
$$\frac{d}{dt} T_t = T_t \circ A = A \circ T_t$$

(Sketch of proof).

□

Outline

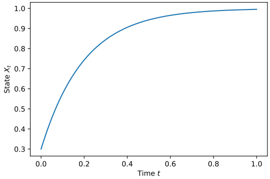
We compare a stochastic system and a fluid approximation



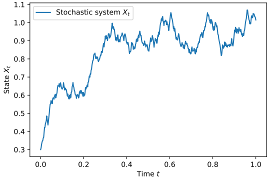
Important notations:

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

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



$\phi_t(X_0)$

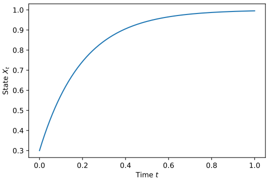


X_t

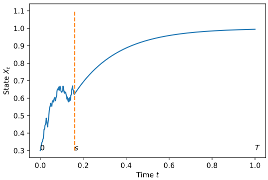
We want to compare:

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

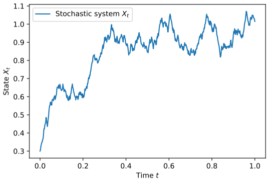
To compare X_t and $\phi_t(X_0)$, we zoom on infinitesimal changes



$\phi_t(X_0)$



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

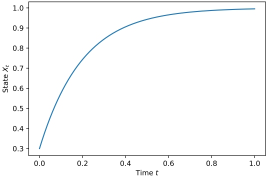


X_t

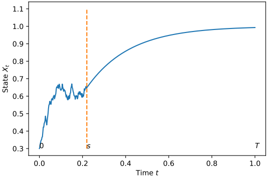
We want to compare:

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

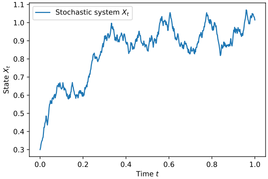
To compare X_t and $\phi_t(X_0)$, we zoom on infinitesimal changes



$\phi_t(X_0)$



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



X_t

We want to compare:

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

Rewriting in terms of generators

$$T_t - \phi_t = \int_0^t \frac{d}{ds} T_s \circ \phi_{t-s} ds$$

$$=$$

Rewriting in terms of generators

$$
\begin{aligned}
T_t - \phi_t &= \int_0^t \frac{d}{ds} T_s \circ \phi_{t-s} ds \\
&= \int_0^t \left(T_s \circ \frac{d}{ds} \phi_{t-s} + ds(T_s) \circ \phi_{t-s} \right) ds \\
&= \int_0^t \left(T_s \circ A^{\mathrm{sto}} \circ \phi_{t-s} - T_s \circ A^{\mathrm{ODE}} \circ \phi_{t-s} \right) ds \\
&= \int_0^t T_s \circ \left(A^{\mathrm{sto}} - A^{\mathrm{ODE}} \right) \circ \phi_{t-s} ds
\end{aligned}
$$

Rewriting in terms of generators (continued)

By using that

$$T_s \circ \left(A^{\mathrm{sto}} - A^{\mathrm{ODE}}\right) \circ \phi_{t-s}(x) = \mathbb{E}\left[\left(A^{\mathrm{sto}} - A^{\mathrm{ODE}}\right) \circ \phi_{t-s}(X_s) \mid X_0 = x\right],$$

the previous slide rewrites as:

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

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

Rewriting in terms of generators (continued)

By using that

$$T_s \circ \left(A^{\mathrm{sto}} - A^{\mathrm{ODE}}\right) \circ \phi_{t-s}(x) = \mathbb{E}\left[\left(A^{\mathrm{sto}} - A^{\mathrm{ODE}}\right) \circ \phi_{t-s}(X_s) \mid X_0 = x\right],$$

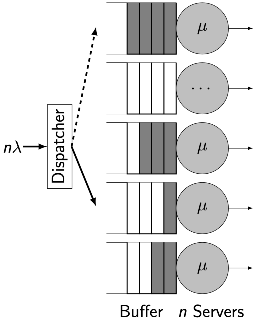
the previous slide rewrites as:

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

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

Outline

Classical Mean Field Setting



Example: Load-balancing

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

Example: the supermarket model (JSQ(d))

$X_i(t) = \{\text{fraction of queues with } i \text{ or more jobs}\}$

Transitions:

1. Departure of a job from a queue of size i:

2. Arrival of a job in a queue of size i:

Example: the supermarket model (JSQ(d))

$X_i(t) = \{\text{fraction of queues with } i \text{ or more jobs}\}$

Transitions:

1. Departure of a job from a queue of size i:

$$X_i \to X_i - \frac{1}{n}\mathbf{e}_i \qquad \text{at rate } n\mu(X_{i-1} - X_i)$$

2. Arrival of a job in a queue of size i:

$$X_i \to X_i + \frac{1}{n}\mathbf{e}_i \qquad \text{at rate } n\lambda(X_{i+1}^d - X_i^d)$$

The associated ODE is:

Example: the supermarket model (JSQ(d))

$X_i(t) = \{\text{fraction of queues with } i \text{ or more jobs}\}$

Transitions:

1. Departure of a job from a queue of size i:

$$X_i \to X_i - \frac{1}{n}\mathbf{e}_i \qquad \text{at rate } n\mu(X_{i-1} - X_i)$$

2. Arrival of a job in a queue of size i:

$$X_i \to X_i + \frac{1}{n}\mathbf{e}_i \qquad \text{at rate } n\lambda(X_{i+1}^d - X_i^d)$$

The associated ODE is:

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

It has a unique fixed point that is an attractor.

Mean field approximation and result

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:

$$\mathbb{E}[X_t] = \phi_t(X_0) + \frac{1}{n}V(t) + O(\frac{1}{n^2}).$$

(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 values of n

	Coupon	Supermarket	Pull/push
Simulation ($N = 10$)	1.530	2.804	2.304
Refined mean field ($N = 10$)	1.517	2.751	2.295
Mean field ($N = \infty$)	1.250	2.353	1.636

Table: Table from "A refined mean field approximation" (G. Van Houdt Sigmetrics 2018)

where

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

Proof (1/2) Generator of a density dependent population process

The generator of the density dependent population process is:

$$A^{\mathrm{sto}}h(x) =$$

Proof (1/2) Generator of a density dependent population process

The generator of the density dependent population process is:

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

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

Proof (1/2) Generator of a density dependent population process

The generator of the density dependent population process is:

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

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

As a consequence

- $(A^{\mathrm{sto}} - A^{\mathrm{ODE}})h = O(1/n)$ if h is C^2.
- The hidden constant depends on $\left\| \nabla^2 h \cdot Q \right\|$.

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

Recall that

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

The function $h(x) = \phi_s(x)$ is C^2 if the drift function f is C^2.

This function belongs to "$\mathcal{H}$" $\Rightarrow$ Error $= O(1/n)$.

Outline

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

Recall that

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

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

Recall that

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

We can take $t \to \infty$ if X_t has a stationary distribution:

$$\mathbb{E}\left[X_\infty\right] - \phi_\infty(X_0) = \mathbb{E}\left[\int_0^\infty (A^{\mathrm{ODE}} - A^{\mathrm{sto}})\phi_s(X_\infty)ds\right]$$
$$= \mathbb{E}\left[(A^{\mathrm{ODE}} - A^{\mathrm{sto}}) \int_0^\infty \phi_s(X_\infty)ds\right]$$

(provided that the above make sense)

How to define $(A^{\mathrm{ODE}} - A^{\mathrm{sto}}) \int_0^\infty \phi_s(x) ds$?

- Fact: If f is a constant function, then $Af = 0$.

Hence, for any constant c:

$$(A^{\mathrm{ODE}} - A^{\mathrm{sto}}) \int_0^t \phi_s(x) ds = (A^{\mathrm{ODE}} - A^{\mathrm{sto}}) \int_0^t (\phi_s(x) - c) ds.$$

How to define $(A^{\mathrm{ODE}} - A^{\mathrm{sto}}) \int_0^\infty \phi_s(x) ds$?

- Fact: If f is a constant function, then $Af = 0$.

Hence, for any constant c:
$$(A^{\mathrm{ODE}} - A^{\mathrm{sto}}) \int_0^t \phi_s(x) ds = (A^{\mathrm{ODE}} - A^{\mathrm{sto}}) \int_0^t (\phi_s(x) - c) ds.$$

Assume that the ODE has a unique attractor x^* that is exponentially stable and take $c = x^*$. We are looking at:

$$(A^{\mathrm{ODE}} - A^{\mathrm{sto}}) \underbrace{\int_0^\infty (\phi_s(x) - x^*) ds}_{\text{Solution of a Poisson equation.}} .$$

Zoom on the Poisson equation

Let G be a generator that has a stationary distribution π. For a function h, we denote by $\bar{h} = \sum_x h(x)\pi(x)$.

- A solution of the poisson equation is a function P_h such that

$$AP_h(x) = \bar{h} - h(x). \qquad (1)$$

  - Solution to (1) is not unique in general.
  - Unique up to additive constant if the process is unichain.

Zoom on the Poisson equation

Let G be a generator that has a stationary distribution π. For a function h, we denote by $\bar{h} = \sum_x h(x)\pi(x)$.

- A solution of the poisson equation is a function P_h such that

$$AP_h(x) = \bar{h} - h(x). \tag{1}$$

  - Solution to (1) is not unique in general.
  - Unique up to additive constant if the process is unichain.
- $P_h(x)$ represents how far x is from the stationary distribution. Indeed, if X_t is a system whose generator is A, a solution to this equation is:

$$P_h(x) = \int_0^\infty (h(X_t) - \bar{h})dt.$$

A solution of a Poisson equation is a bias

Link with MDPs and Markov reward process
Consider a Markov chain (X_t) and assume that you earn a reward h. Then:

$$\int_0^T \mathbb{E}\left[h(X_t)\right] dt = T\bar{h} + \underbrace{P_h(X_0)}_{\text{bias}} + o(1).$$

If the Markov chain has a transition kernel K, then P_h satisfies:

$$P_h(x) - \bar{h} = h(x) + \sum_y K_{xy} P_h(y).$$

Poisson equation and ODE

For the ODE, let P_{Id} be defined as:

$$P_h(x) = \int_0^\infty h(\phi_s(x)) - h(x^*) ds.$$

We have $A^{\mathrm{ODE}} P_h(x) = h(x^*) - h(x)$.

Poisson equation and ODE

For the ODE, let P_{Id} be defined as:

$$P_h(x) = \int_0^{\infty} h(\phi_s(x)) - h(x^*) ds.$$

We have $A^{\mathrm{ODE}} P_h(x) = h(x^*) - h(x)$.

We can use this to retrieve Stein's method:

$$\begin{aligned} \mathbb{E}\left[h(x^*) - h(X_\infty)\right] &= \mathbb{E}\left[A^{\mathrm{ODE}} P_h(X_\infty)\right] \\ &= \mathbb{E}\left[(A^{\mathrm{ODE}} - A^{\mathrm{sto}}) P_h(X_\infty)\right] \end{aligned}$$

Consequence for the error of fluid approximation

Recall that for finite t:

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

Consequence for the error of fluid approximation

Recall that for finite t:

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

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

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

Consequence for the error of fluid approximation

Recall that for finite t:

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

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

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

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

- for some $h \in \mathcal{H}$, $(A^{\text{ODE}} - A^{\text{sto}})h$ is small.
- ϕ_{t-s} and P_h belong to this $\mathcal{H}$.

Application to density dependent population processes

If the ODE $\dot{x} = f(x)$ has an exponentially stable attractor and if f is C^2.
Then:

$$x \mapsto \int_0^\infty (\phi_s(x) - x^*) ds \text{ is } C^2.$$

Consequence: if a density dependent population process has a unique stable attractor, then

$$\mathbb{E}[X_\infty] = x^* + \frac{1}{n} V + O(\frac{1}{n^2}).$$

Application to density dependent population processes

If the ODE $\dot{x} = f(x)$ has an exponentially stable attractor and if f is C^2.
Then:

$$x \mapsto \int_0^\infty (\phi_s(x) - x^*) ds \text{ is } C^2.$$

Consequence: if a density dependent population process has a unique stable attractor, then

$$\mathbb{E}[X_\infty] = x^* + \frac{1}{n}V + O(\frac{1}{n^2}).$$

	Coupon	Supermarket	Pull/push
Simulation ($N = 10$)	1.530	2.804	2.304
Refined mean field ($N = 10$)	1.517	2.751	2.295
Mean field ($N = \infty$)	1.250	2.353	1.636

Figure: Table from "A refined mean field approximation" (G. Van Houdt Sigmetrics 2018)

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.
- G. Allmeier 2024: Extension to 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: optimizaton (bandit problems), (sparse) geometric models, non-Markovian.

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.
- 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

Outline

What is a two-timescale stochastic process?

Slow process $X \in \mathbb{R}^{d_x}$	Fast process $Y \in \{1 \dots d_y\}$
X jumps from x to $x + \frac{1}{n}\ell$ • at rate $n\beta_\ell(x, y)$.	Y jumps from y to y' • at rate $nK^{fast}_{y,y'}(x)$.

What is a two-timescale stochastic process?

Slow process $X \in \mathbb{R}^{d_x}$	Fast process $Y \in \{1 \dots d_y\}$
X jumps from x to $x + \frac{1}{n}\ell$ • at rate $n\beta_\ell(x, y)$.	Y jumps from y to y' • at rate $nK^{fast}_{y,y'}(x)$.

Two approximations:

Drift $\dot{X} \approx f(X, Y)$	$\mathbf{P}[Y(t) = y] \approx \pi_y(X(t))$

What is a two-timescale stochastic process?

Slow process $X \in \mathbb{R}^{d_x}$	Fast process $Y \in \{1 \dots d_y\}$
X jumps from x to $x + \frac{1}{n}\ell$ • at rate $n\beta_\ell(x, y)$.	Y jumps from y to y' • at rate $nK^{fast}_{y,y'}(x)$.

Two approximations:

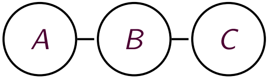
Drift $\dot{X} \approx f(X, Y)$	$\mathbf{P}\left[Y(t) = y\right] \approx \pi_y(X(t))$

The ODE is $\dot{x} = \sum_y \pi_y(x) f(x, y)$

(Averaging technique)

Example: CSMA with queues

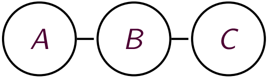
Model from Cecchi, Borst, Leeuwaarden 2015.



Interference graph, n nodes per class A, B or C.

Example: CSMA with queues

Model from Cecchi, Borst, Leeuwaarden 2015.



Interference graph, n nodes per class A, B or C.

"Slow process": X.

$X_{i,s}$ = proportion of nodes of class i with $\geq s$ messages

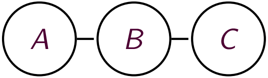
Arrival/departure:

$$X_{i,s} \mapsto X_{i,s} \pm \frac{1}{n}$$

Rate depends on Y.

Example: CSMA with queues

Model from Cecchi, Borst, Leeuwaarden 2015.



Interference graph, n nodes per class A, B or C.

"Slow process": X.

$X_{i,s}$ = proportion of nodes of class i with $\geq s$ messages

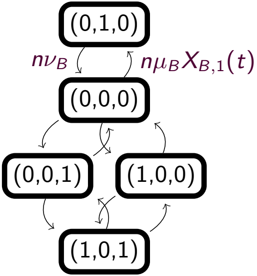
Arrival/departure:

$$X_{i,s} \mapsto X_{i,s} \pm \frac{1}{n}$$

Rate depends on Y.

"Fast process": Y.

$Y_i = 1$ if class i talks.



Accuracy results (Allmeier, G. 2022)

We consider a "generic" multiscale model as in Slide 3.

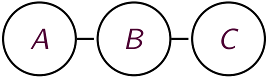
Theorem. Assume that $f(X, Y)$ and $K(X)$ are twice differentiable in X, and that $K^{fast}(X)$ is "unichain" for all X, then:

$$\mathbb{E}[X(t)] = \phi_t(X_0) + \frac{1}{n}V(t) + O(1/n^2).$$

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

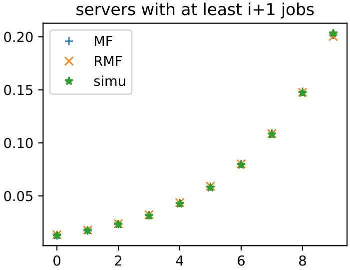
Again, the refined approximation is very accurate

Example with $n = 1$ node per class.



Jobs arrive at rate 1, activation rate $= 3$. Job duration is $1/3$.

Class A or C



Class B

Proof (1/3): comparison of generator is not sufficient!

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

$$\begin{aligned}
A^{\text{sto}} h(X, Y) &= \sum_{\ell} (h(X + \frac{1}{n}\ell) - h(X)) n\beta_\ell(X, Y) \\
&= \nabla h \cdot f(X, Y) + O(1/n^2). \\
A^{\text{ODE}} h(X, Y) &= \nabla h \cdot \sum_{y} \pi_y(X) f(X, Y).
\end{aligned}$$

Proof (1/3): comparison of generator is not sufficient!

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

$$
\begin{aligned}
A^{\mathrm{sto}} h(X,Y) &= \sum_{\ell} (h(X + \frac{1}{n}\ell) - h(X)) n\beta_\ell(X,Y) \\
&= \nabla h \cdot f(X,Y) + O(1/n^2). \\
A^{\mathrm{ODE}} h(X,Y) &= \nabla h \cdot \sum_y \pi_y(X) f(X,Y).
\end{aligned}
$$

Hence:

$$
(A^{\mathrm{sto}} - A^{\mathrm{ODE}}) h(X,Y) = \nabla h \cdot \underbrace{\Big(f(X,Y) - \sum_y \pi_y(X) f(X,Y) \Big)}_{\text{this was } = 0 \text{ for the singlescale model.}} + O(1/n)
$$

Proof (2/3): Introducing the Poisson equation for the fast system

Let P_f^{fast} the solution of the Poisson equation for $K^{fast}(x)$ and f. Then:

$$f(X,Y) - \sum_y \pi_y(X) f(X,Y) = K^{fast}(x) P_f^{fast}(X,Y).$$

Lemma: if $K(x)$ is C^2 and unichain for all x, then P_f^{fast} is C^2.

Proof (3/3)

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

$$A^{\mathrm{sto}} h(X, Y) = n K^{fast} h(X, Y) + \nabla_x h \cdot f(X, Y) + O(1/n)$$

Hence, $K^{fast} = \frac{1}{n} A^{\mathrm{sto}}$ if h is C^1.

Proof (3/3)

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

$$A^{\mathrm{sto}} h(X, Y) = nK^{fast} h(X, Y) + \nabla_x h \cdot f(X, Y) + O(1/n)$$

Hence, $K^{fast} = \frac{1}{n} A^{\mathrm{sto}}$ if h is C^1.

Going back to the previous slides, we have:

$$(A^{\mathrm{sto}} - A^{\mathrm{ODE}}) h(X, Y) = K^{fast} \nabla h \cdot P_f^{fast}(X, Y) + O(1/n).$$

Hence, we are left with terms of the form

$$\begin{aligned}\int_0^s K^{fast} P_f^{fast}(X_s, Y_s) ds &= \int_0^s \frac{1}{n} A^{\mathrm{sto}} P_f^{fast}(X_s, Y_s) ds \\ &= \frac{1}{n}(P_f^{fast}(X_t, Y_t) - P_f^{fast}(X_0, Y_0)) \\ &= O(1/n).\end{aligned}$$

The last is because P_f^{fast} is C^1 when f and K are C^1.

Recap on multi-scale

This shows that the $O(1/n)$-expansion also holds for multiscale model provided that:

- Transitions are C^2 (as always).
- $K(x)$ is unichain for all x.

We can compute the expansion-term for $t = +\infty$.