Refinements of Mean Field Approximation

HDR defense – Nicolas Gast

Inria

January 30, 2020. Grenoble

The objective of my work is to provide tools to:

- Describe distributed systems;
- Optimize their behavior.

Good system design needs performance evaluation.

Tools:

- Stochastic modeling, Markov Chains.
- Dynamical systems.
- Optimization, optimal control.

Example 1: Networking and Congestion Control



¹ "MPTCP is not pareto-optimal: performance issues and a possible solution". (CoNext 2012 (best paper), ToN 2013) by Khalili, Gast, Popovic, and Le Boudec.



 $^{^{1}}$ "MPTCP is not pareto-optimal: performance issues and a possible solution". (CoNext 2012 (best paper), ToN 2013) by Khalili, Gast, Popovic, and Le Boudec.

3/29

Example 2: Load balancing



N servers

Which allocation policy?

- Random;
- Round-robin;
- *JSQ*;
- *JSQ*(*d*);
- JIQ;

• . . .

4 / 29

² A mean field model of work stealing in large-scale systems (Sigmetrics 2010) by Gast and Gaujal, A refined mean field approximation (Sigmetrics 2018 (best paper)) by Gast and Van Houdt

Example 2: Load balancing



Which allocation policy?

- Random;
- Round-robin;
- *JSQ*;
- *JSQ*(*d*);
- JIQ;
- . . .

² A mean field model of work stealing in large-scale systems (Sigmetrics 2010) by Gast and Gaujal, A refined mean field approximation (Sigmetrics 2018 (best paper)) by Gast and Van Houdt

This leads us to study markovian models of interacting agents.



This leads us to study markovian models of interacting agents.



Problem: state space explosion. S states per object, N objects $\Rightarrow S^N$ states. In my work, I develop and use models that scale.

Today, I will mostly focus on mean field approximation.





Nicolas Gast (Inria)







Idea of mean field approximation

Each individual interacts with the mass while having a negligible effect on the mass.



- Mean field theory (Statistical mechanics, 1800s-1900s)
- Theoretical biology
- Computer modeling (ex: Baccelli 92, Vvedenskaya 96)
- Mean field games (Lasry-Lions 2007)

Idea of mean field approximation

Each individual interacts with the mass while having a negligible effect on the mass.



- Mean field theory (Statistical mechanics, 1800s-1900s)
- Theoretical biology
- Computer modeling (ex: Baccelli 92, Vvedenskaya 96)
- Mean field games (Lasry-Lions 2007)

Example: The supermarket model (SQ(2))



Randomly choose two, select one.

N identical queues.

- Arrival rate $N\lambda$;
- Service rate μ .

State = queue sizes.

Q(t) = (1, 1, 2, 3).

Example: The supermarket model (SQ(2))



N identical queues.

- Arrival rate $N\lambda$;
- Service rate μ .

Q(t) = (1, 1, 2, 3).

If all queues are identical, we can simplify: $X = (X_1, X_2, ...)$

 $X_i(t)$ = fraction of queues with queue length $\geq i$.

Above: X(t) = (1, .5, .25, 0, 0, 0, ...).

SQ(2): Transitions and Mean Field Approximation.

The fractions of queue with *i* jobs or more changes when:

A job arrives:
$$X_i \mapsto X_i + \frac{1}{N}$$
 (at rate $N\lambda(X_{i-1}^2 - X_i^2)$)
A job departs: $X_i \mapsto X_i - \frac{1}{N}$ (at rate $N\mu(X_i - X_{i+1})$).

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

$$\dot{x}_i = \underbrace{\lambda(x_{i-1}^2 - x_i^2)}_{\text{Arrival}} - \underbrace{\mu(x_i - x_{i+1})}_{\text{Departure}}$$

SQ(2): Transitions and Mean Field Approximation.

The fractions of queue with *i* jobs or more changes when:

Nicolas Gast (Inria)

A job arrives:
$$X_i \mapsto X_i + \frac{1}{N}$$
 (at rate $V\lambda(X_{i-1}^2 - X_i^2)$)
A job departs $X_i \mapsto X_i - \frac{1}{N}$ (at rate $V\mu(X_i - X_{i+1})$).
 $x \mapsto x + \frac{1}{N}\ell$ Rate $Nr(\ell, x)$

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

$$\begin{aligned}
\hat{x}_{i} &= \underbrace{\lambda(x_{i-1}^{2} - x_{i}^{2})}_{\text{Arrival}} - \underbrace{\mu(x_{i} - x_{i+1})}_{\text{Departure}} \\
\text{Drift } f(x) &= \sum_{\ell} \ell r(\ell, x).
\end{aligned}$$
This is a density dependent population process.
$$\begin{aligned}
\text{Nicolas Gast (Inria)} &= \underbrace{Refinements of Mean Field Approximation}_{\text{January 30, 2020, Grenoble}} &= 10/29
\end{aligned}$$

Refinements of Mean Field Approximation

Mean field is asymptotically exact



Mean field is asymptotically exact



The stochastic approximation approach:

$$X\left(t+rac{1}{N}
ight)=X(t)+rac{1}{N}\Big(f(X(t))+ ext{noise}\Big).$$

Mean field is asymptotically exact



The stochastic approximation approach:

$$X\left(t+rac{1}{N}
ight)=X(t)+rac{1}{N}\Big(f(X(t))+ ext{noise}\Big).$$

Nicolas Gast (Inria)

A few contributions

Mean field approximation is exact as $N \to \infty$. We extended this methodology to . . .

- ... study discontinuous or imprecise systems;
 - Differential inclusions
 - Numerical algorithms.
- ... simplify optimal control problems.
 - Discrete time
 - Continuous time
 - Mean field game

[PEVA 2012, DSN 2016] [QEST 2019]

> DEDS 2011] [TAC 2012] [JDG 2019]

How large should

Nicolas Gast (Inria)

A few contributions

Mean field approximation is exact as $N \to \infty$. We extended this methodology to . . .

- ... study discontinuous or imprecise systems;
 - Differential inclusions
 - Numerical algorithms.
- ... simplify optimal control problems.
 - Discrete time
 - Continuous time
 - Mean field game

[PEVA 2012, DSN 2016] [QEST 2019]

> [DEDS 2011] [TAC 2012] [JDG 2019]

How large should N be?

Nicolas Gast (Inria)

A few contributions

Mean field approximation is exact as $N \to \infty$. We extended this methodology to . . .

- ... study discontinuous or imprecise systems;
 - Differential inclusions
 - Numerical algorithms.
- ... simplify optimal control problems.
 - Discrete time
 - Continuous time
 - Mean field game

[PEVA 2012, DSN 2016] [QEST 2019]

> [DEDS 2011] [TAC 2012] [JDG 2019]

Key question

How large should N be?

Outline



Mean field approximation is a law of large number

Theorem (Kurtz 1970s... Ying 2016) For homogeneous and regular models

$$\mathbb{E}\left[\|X(t) - x(t)\|\right] = O(1/\sqrt{N})$$

where x is the solution of the ODE $\dot{x} = f(x)$ and f is the drift. (Valid for t > 0 and $t = +\infty$ if exponentially stable attractor.)

Mean field approximation is a law of large number

Theorem (Kurtz 1970s... Ying 2016) For homogeneous and regular models

$$\mathbb{E}\left[\|X(t) - x(t)\|\right] = O(1/\sqrt{N})$$

where x is the solution of the ODE $\dot{x} = f(x)$ and f is the drift. (Valid for t > 0 and $t = +\infty$ if exponentially stable attractor.)



In practice, mean field works well even for "small" systems $_{\rm Why?}$

N	10	100	1000	$+\infty$
Average queue length for $SQ(2)$	3.81	3.39	3.36	3.35

In practice, mean field works well even for "small" systems $_{\rm Why?}$

N	10	100	1000	$+\infty$
Average queue length for $SQ(2)$	3.81	3.39	3.36	3.35



In practice, mean field works well even for "small" systems $_{\rm Why?}$



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

Theorem (Kolokoltsov 2012, G. 2017, G. and Van Houdt 2018). For a density dependent population process X, i f the drift f is twice differentiable, then for any t > 0:

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

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

2 V(t) can be easily computed numerically If it has a unique exponentially stable attractor, this is true for $t = \infty$.

Refined mean field is designed to study finite system



Refined mean field is designed to study finite system



Our results on refined mean field methods

- Mean field is 1/N-accurate
 - ► Stein's method (inspired by Braverman et al 2015 / Ying 2016)
- Refinement for steady-state
 - It is very accurate even for N = 10.
 - We can quantify 1/N variants such as choosing with/without replacement.
- Extension to finite-time and $1/N^2$ approximation
- [Performance 2018]
- Moment-closure approach, tensor decomposition
- Numerical tool⁴

• Synchronous systems

[PEVA 2019]

Nicolas Gast (Inria)

[Sigmetrics 2017]

best paper award

^{&#}x27;https://github.com/ngast/rmf_tool

Our results on refined mean field methods

- Mean field is 1/N-accurate
 - Stein's method (inspired by Braverman et al 2015 / Ying 2016)
- Refinement for steady-state
 - It is very accurate even for N = 10.
 - We can quantify 1/N variants such as choosing with/without replacement.
- Extension to finite-time and $1/N^2$ approximation [Performance 2018]

- Moment-closure approach, tensor decomposition
- ▶ Numerical tool⁴

Synchronous systems

³best paper award

Nicolas Gast (Inria)

[Sigmetrics 2017]

[Sigmetrics 2018³]
Our results on refined mean field methods

- Mean field is 1/N-accurate
 - Stein's method (inspired by Braverman et al 2015 / Ying 2016)
- Refinement for steady-state
 - It is very accurate even for N = 10.
 - We can quantify 1/N variants such as choosing with/without replacement.
- Extension to finite-time and $1/N^2$ approximation

[Performance 2018]

- Moment-closure approach, tensor decomposition
- Numerical tool⁴

• Synchronous systems

[PEVA 2019]

³best paper award ⁴https://github.com/ngast/rmf_tool

Nicolas Gast (Inria)

[Sigmetrics 2017] 2016)

[Sigmetrics 2018³]

Our results on refined mean field methods

- Mean field is 1/*N*-accurate
 - Stein's method (inspired by Braverman et al 2015 / Ying 2016)
- Refinement for steady-state
 - It is very accurate even for N = 10.
 - We can quantify 1/N variants such as choosing with/without replacement.
- Extension to finite-time and $1/N^2$ approximation
- [Performance 2018]
- Moment-closure approach, tensor decomposition
- Numerical tool⁴

Synchronous systems

[PEVA 2019]

[Sigmetrics 2017]

[Sigmetrics 2018³]

³best paper award

⁴https://github.com/ngast/rmf_tool

Our results on refined mean field methods

- Mean field is 1/N-accurate
 - Stein's method (inspired by Braverman et al 2015 / Ying 2016)
- Refinement for steady-state
 - It is very accurate even for N = 10.
 - We can quantify 1/N variants such as choosing with/without replacement.
- Extension to finite-time and $1/N^2$ approximation

[Performance 2018]

- Moment-closure approach, tensor decomposition
- Numerical tool⁴

Synchronous systems

[PEVA 2019]

Nicolas Gast (Inria)

[Sigmetrics 2018³]

[Sigmetrics 2017]

³best paper award

⁴https://github.com/ngast/rmf_tool

The approximation is very accurate, even for N = 10. Average queue length for supermarket SQ(2) model.

Ν	10	20	30	50	100	∞
ho = 0.9						
Simulation ("exact")	2.804	2.567	2.491	2.434	2.393	_
Refined mean field	2.751	2.552	2.486	2.432	2.393	2.353
Error	0.053	0.015	0.005	0.002	< 0.001	

The approximation is very accurate, even for N = 10. Average queue length for supermarket SQ(2) model.

Ν	10	20	30	50	100	∞	
$\rho = 0.9$ Simulation ("exact")	2 804	2 567	2 /01	Vean fi	eld appro	ximatio	n
Refined mean field	2.751	2.552	2.486	2.432	2.393	2.353	
Error	0.053	0.015	0.005	0.002	< 0.001		1

The accuracy of the classical mean field degrades as ρ approaches 1:

average queue length
$$\approx \log_2 \frac{1}{1-\rho} + \frac{1}{N} \frac{\rho^2}{2(1-\rho)} + \frac{1}{N^2} \frac{1}{20(1-\rho)^2}$$

The moment closure approach

Consider a system for which X becomes X + 1/N at rate NX^2 . We have:

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

The moment closure approach

Consider a system for which X becomes X + 1/N at rate NX^2 . We have:

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

The moment closure approach

Consider a system for which X becomes X + 1/N at rate NX^2 . We have:

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

 $\approx \mathbb{E}[X]^2$ (mean field approx.)

The moment closure approach

Consider a system for which X becomes X + 1/N at rate NX^2 . We have:

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

The moment closure approach

Consider a system for which X becomes X + 1/N at rate NX^2 . We have:

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

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

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

$$\vdots$$

These 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.

Outline



Analysis of Cache (Re)placement Policies



Analysis of Cache (Re)placement Policies



• Popularity-oblivious policies (LRU, RANDOM)

The RAND(\vec{m}) policy

• RANDOM: exchange the requested item with a random item



The RAND(\vec{m}) policy

• RAND (\vec{m}) : exchange the requested item with an item from next list



Conjecture from the 80s: if popularities do not vary, adding more lists is always better.

RAND(m) model

[Sigmetrics 2015]

- Disprove a 30 year old conjecture (more lists is not always better).
- Prove accuracy of a heterogeneous mean field model.

• Extension LRU(*m*) and *q*-LRU variants

- Asymptotically exact analysis.
- Refined approximation for RAND(*m*) variants [Submitted, 2019]
 - ▶ We can define a refined approximation for an heterogeneous model.

RAND(m) model

[Sigmetrics 2015]

- Disprove a 30 year old conjecture (more lists is not always better).
- Prove accuracy of a heterogeneous mean field model.
- Extension LRU(m) and q-LRU variants

- Asymptotically exact analysis.
- Refined approximation for RAND(m) variants [Submitted, 2019]
 We can define a refined approximation for an heterogeneous model.

RAND(m) model

[Sigmetrics 2015]

- Disprove a 30 year old conjecture (more lists is not always better).
- Prove accuracy of a heterogeneous mean field model.
- Extension LRU(m) and q-LRU variants
 - Asymptotically exact analysis.
- Refined approximation for RAND(m) variants [Submitted, 2019]
 - ► We can define a refined approximation for an heterogeneous model.

RAND(m) model

[Sigmetrics 2015]

- Disprove a 30 year old conjecture (more lists is not always better).
- Prove accuracy of a heterogeneous mean field model.
- Extension LRU(m) and q-LRU variants

- Asymptotically exact analysis.
- Refined approximation for RAND(m) variants [Submitted, 2019]
 - ► We can define a refined approximation for an heterogeneous model.

Heterogeneous mean field approximation Let $H_i(t)$ be the popularity in list *i*.



Heterogeneous mean field approximation Let $H_i(t)$ be the popularity in list *i*.



If $x_{k,i}(t)$ is the probability that item k is in list i at time t, we approximately have:

$$\dot{x}_{k,i}(t) = p_k x_{k,i-1}(t) - \underbrace{\sum_{j}^{p_{j}} p_{j} x_{j,i-1}(t)}_{i} \frac{x_{k,i}(t)}{m_i} + \mathbf{1}_{\{i < h\}} \underbrace{\left(\sum_{j}^{p_{j}} p_{j} x_{j,i}(t) \frac{x_{k,i+1}(t)}{m_{i+1}} - p_k x_{k,i}(t)\right)}_{\text{Populatity in cache i}}$$

This approximation is of the form $\dot{x} = xQ(x)$.

The mean field ODE is asymptotically exact

"Theorem": For any T, there exists C such that for any popularities p and any list sizes $m_1 \dots m_h$



1000 items (Zipf). Popularities change every 2000 requests

Nicolas Gast (Inria)

We can refine the approximation

- We define a scaled process with N copies of each item.
- We use $Perf(N) \approx Perf(\infty) + \frac{V}{N}$ with N = 1.

We can refine the approximation

- We define a scaled process with N copies of each item.
- We use $Perf(N) \approx Perf(\infty) + \frac{V}{N}$ with N = 1.

Future research directions:

Online optimization algorithms

Future research directions:

 Online optimization algorithms

Nicolas Gast (Inria)

Thank you

Advisors

Former PhD students / post-docs

The refined approximation can also account for behaviors that are indistinguishable by classical mean field methods Example: choosing with or without replacement

Let x_i be the fraction of queues of size *i* or more. Pick two queues, what is the probability that smallest has size *i*?

• If picked with replacement: $x_i^2 - x_{i+1}^2$. • If picked without replacement: $x_i \frac{Nx_i - 1}{N - 1} - x_{i+1} \frac{Nx_{i+1} - 1}{N - 1}$

The refined approximation can also account for behaviors that are indistinguishable by classical mean field methods Example: choosing with or without replacement

Let x_i be the fraction of queues of size *i* or more. Pick two queues, what is the probability that smallest has size *i*?

 \bigcirc

• If picked with replacement: $x_i^2 - x_{i+1}^2$. • If picked without replacement: $x_i \frac{Nx_i - 1}{N - 1} - x_{i+1} \frac{Nx_{i+1} - 1}{N - 1}$ For $\rho = 0.9$: Average queue length (with rep) $\approx 2.353 + \frac{4}{N}$ Average queue length (without rep) $\approx 2.353 + \frac{4}{N} - \frac{1}{N}$ Nicolas Gast (Intia) Refinements of Mean Field Approximation January 30, 2020. Grenoble 1/3

The refined approximation can also account for behaviors that are indistinguishable by classical mean field methods Example: choosing with or without replacement

Let x_i be the fraction of queues of size *i* or more. Pick two queues, what is the probability that smallest has size *i*?

How does the expected queue length evolve with time?

2/3

How does the expected queue length evolve with time?

How does the expected queue length evolve with time?



Remark about computation time :

- 10min/1h (simulation N = 1000/N = 10), C++ code. Requires many simulations, confidence intervals,...
- 80ms (mean field), 700ms (1/N-expansion), 9s ($1/N^2$ -expansion), Python numpy

Does it always work?

Can I always exchange the limits $N \to \infty$, $k \to \infty$, $t \to \infty$?

$$\mathbb{E}[X(t)] = x(t) + \frac{1}{N}V(t) + \frac{1}{N^2}A(t) + \dots + O(\frac{1}{N^{k+1}})$$

Does it always work?

Can I always exchange the limits $N \to \infty$, $k \to \infty$, $t \to \infty$?

$$\mathbb{E}[X(t)] = x(t) + \frac{1}{N}V(t) + \frac{1}{N^2}A(t) + \dots + O(\frac{1}{N^{k+1}})$$



Nicolas Gast (Inria)