Expected Values Estimated via Mean-Field Approximation are $1/N$-Accurate

Nicolas Gast

Inria, Grenoble, France

Sigmetrics 2017, Urbana-Champaign

Mean-field approximation is widely used in our community

2016 **Asymptotics of Insensitive Load Balancing and Blocking Phases** – Jonckheere - Prabhu

2016 **On the Approximation Error of Mean-Field Models** – Ying

2015 **Power of d Choices for Large-Scale Bin Packing: A Loss Model** – Xie et al

2015 **Transient and Steady-state Regime of a Family of List-based Cache Replacement Algorithms** – Gast, Van Houdt

2013 **Queueing system topologies with limited flexibility.** – Tsitsiklis, Xu

2013 **A mean field model for a class of garbage collection algorithms in flash-based solid state drives.** – Van Houdt

2012 **Fluid limit of an asynchronous optical packet switch with shared per link full range wavelength conversion.** – Van Houdt, Bortolussi

2011 **On the power of (even a little) centralization in distributed processing.** –

2010 **Randomized load balancing with general service time distributions.** – Bramson et al.

2010 **Incentivizing peer-assisted services: a fluid shapley value approach.** – Misra et al

2010 **A mean field model of work stealing in large-scale systems.** – Gast, Gaujal

2009 **The age of gossip: spatial mean field regime.** – Chaintreau et al.

⋮

What is mean-field approximation ?

We study a population of N interchangeable objects.

$X^{(N)}$ denotes the empirical measure.

$$X_i^{(N)}(t) = \text{ fraction of objects in state } i$$

Idea of mean-field: Some models simplify as $N \to \infty$

Theorem (Kurtz 70,...)

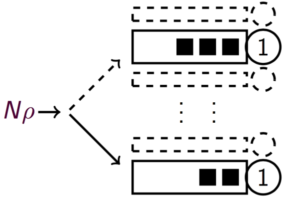
$$X^{(N)}(t) \approx x(t)$$

Idea of mean-field: Some models simplify as $N \to \infty$

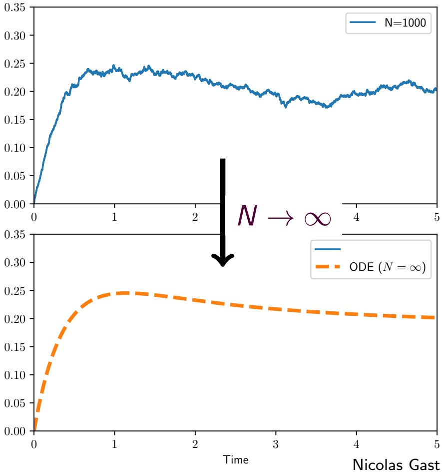
Theorem (Kurtz 70,...)

$$X^{(N)}(t) \approx x(t)$$

Example: N servers



Randomly choose two, and select one

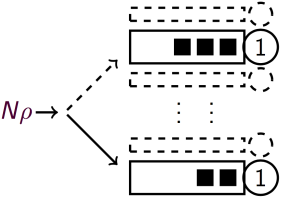


Idea of mean-field: Some models simplify as $N \to \infty$

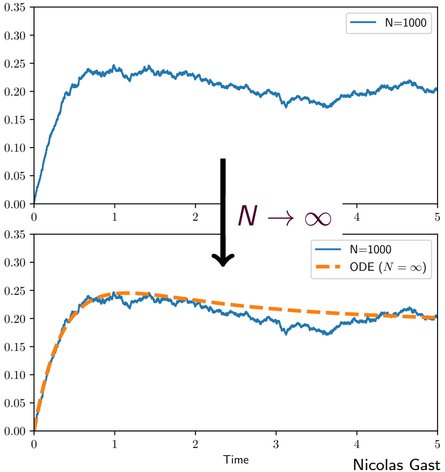
Theorem (Kurtz 70,...)

$$X^{(N)}(t) \approx x(t)$$

Example: N servers



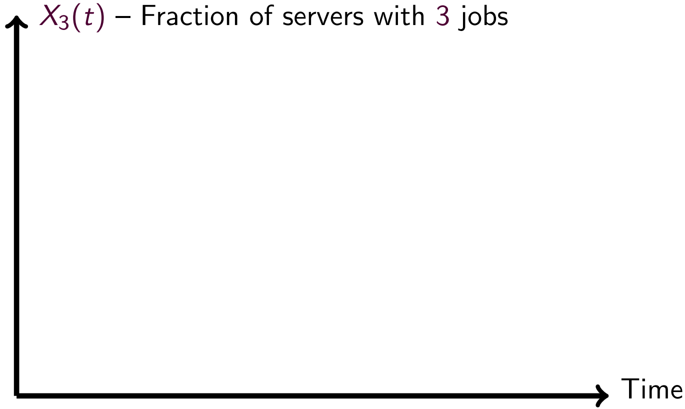
Randomly choose two, and select one



How accurate is mean-field approximation?

Theorem (Kurtz 70...)

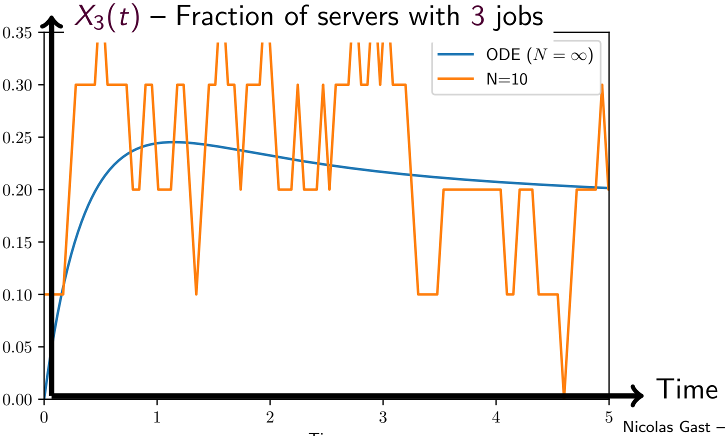
$$X^{(N)}(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$



How accurate is mean-field approximation?

Theorem (Kurtz 70...)

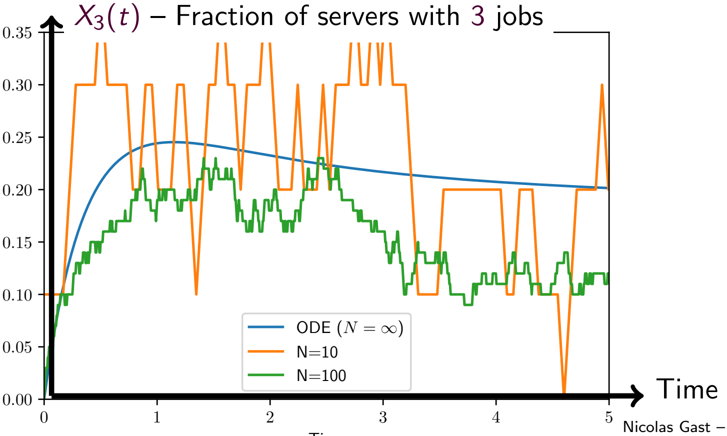
$$X^{(N)}(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$



How accurate is mean-field approximation?

Theorem (Kurtz 70... Ying 16)

$$X^{(N)}(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$

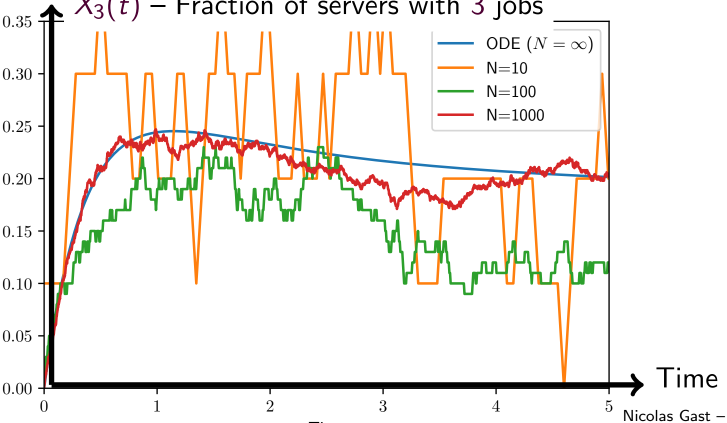


How accurate is mean-field approximation?

Theorem (Kurtz 70... Ying 16)

$$X^{(N)}(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$

$X_3(t)$ – Fraction of servers with 3 jobs



In practice, how accurate is mean-field approximation?

Can we use the approximation for $N = 1000$? $N = 100$? $N = 10$?

N	10	100	1000	∞
Average queue length (simulation)	2.8040	2.3931	2.3567	2.3527

Table: Two-choice model with $\rho = 0.9$

In practice, how accurate is mean-field approximation?

Can we use the approximation for $N = 1000$? $N = 100$? $N = 10$?

N	10	100	1000	∞
Average queue length (simulation)	2.8040	2.3931	2.3567	2.3527
Error of mean-field	0.4513	0.0404	0.0040	0

Table: Two-choice model with $\rho = 0.9$

In practice, how accurate is mean-field approximation?

Can we use the approximation for $N = 1000$? $N = 100$? $N = 10$?

N	10	100	1000	∞
Average queue length (simulation)	2.8040	2.3931	2.3567	2.3527
Error of mean-field	0.4513	0.0404	0.0040	0

Table: Two-choice model with $\rho = 0.9$

Contributions :

1. We show that under very general conditions, the error is in $O(1/N)$
2. We show that for that, the drift needs to be twice-differentiable.
3. We study numerically the power-of-two choice.

Outline

Outline

CTMC

A continous-time Markov chain (CTMC) with state-space **E** is given by an initial state x_0 and its transitions ($\ell \in \mathcal{L}$):

$$X \mapsto X + \ell \qquad \text{at rate} \quad \beta_\ell(X).$$

The drift is $f(x) = \sum_\ell \ell \beta_\ell(x)$.

[1] We assume $(\mathbf{E}, \|\cdot\|)$ is a Banach space, not necessarily $\mathbb{R}^d$.

Population CTMC

Density dependent population process (70s)

A population process is a sequence of CTMC $\mathbf{X}^N$, indexed by the population size N, with state spaces $\mathbf{E}^N \subset \mathbf{E}$, with initial state x_0 and with transitions (for $\ell \in \mathcal{L}$):

$$X \mapsto X + \frac{\ell}{N} \qquad \text{at rate } N\beta_\ell(X).$$

The drift is $f(x) = \sum_\ell \ell\beta_\ell(x)$.

[1] We assume $(\mathbf{E}, \|\cdot\|)$ is a Banach space, not necessarily $\mathbb{R}^d$.

Transient regime

Let Φ_t denotes the (unique) solution of the ODE :

$$\Phi_t x = x + \int_0^t f(\Phi_s x) ds.$$

Transient regime

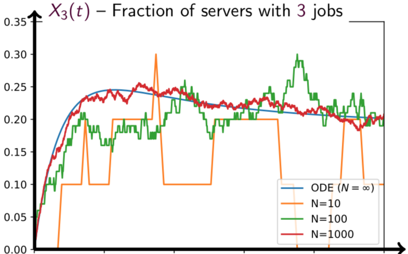
Let Φ_t denotes the (unique) solution of the ODE :

$$\Phi_t x = x + \int_0^t f(\Phi_s x) ds.$$

Theorem (Kurtz 70s)

If f is Lipschitz-continuous with constant L, then for any fixed T:

$$\sup_{t<T} \left\| X^N(t) - \Phi_t X^N(0) \right\| = O(1/\sqrt{N}) \qquad (\lim_{N\to\infty} \cdot = 0).$$



Stationary regime

If the ODE $\dot{x} = f(x)$ has a unique fixed point x^* that is exponentially stable, then:

Theorem (Ying 2016)

If f is Lipschitz-continuous with constant L, then for any fixed T:

$$\mathbb{E}\left[\left\|X^N - x^*\right\|\right] = O(1/\sqrt{N}) \qquad\qquad (\lim_{N\to\infty} \cdot = 0).$$

(the uniqueness of the fixed point is not sufficient, see Benaim-Le Boudec 2008).

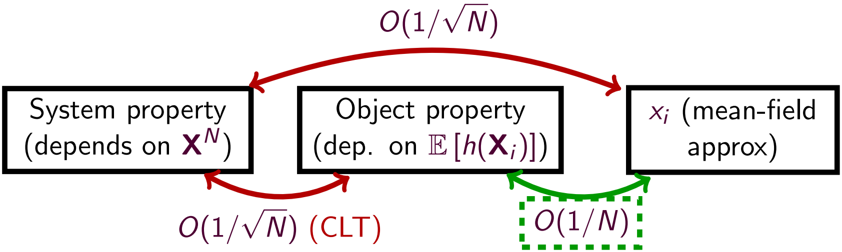
Outline

$1/\sqrt{N}$ or $1/N$?

$$\boxed{\text{System property (depends on } \mathbf{X}^N)} \xleftrightarrow{O(1/\sqrt{N})} \boxed{x_i \text{ (mean-field approx)}}$$

N	10	100	1000	$+\infty$
Average queue length (m^N)	2.81	2.39	2.36	2.35
Error ($m^N - m^\infty$)	0.46	0.039	0.004	0

$1/\sqrt{N}$ or $1/N$?



N	10	100	1000	$+\infty$
Average queue length (m^N)	2.81	2.39	2.36	2.35
Error ($m^N - m^\infty$)	0.46	0.039	0.004	0

Steady-state analysis

We say that $\dot{x} = f(x)$ has an exponentially stable attractor x^* if for any solution:

$$\|x(t) - x^*\| \leq Ce^{-\alpha t} \|x(0) - x^*\| .$$

Steady-state analysis

We say that $\dot{x} = f(x)$ has an exponentially stable attractor x^* if for any solution:

$$\|x(t) - x^*\| \leq Ce^{-\alpha t} \|x(0) - x^*\| .$$

Theorem

If f is twice differentiable, if the ODE has an exponentially stable attractor x^* and if there exists a bounded set $\mathcal{B}$ such that $\mathbf{P}\left[X^N \notin \mathcal{B}\right] = O(1/N^2)$, then for any bounded function h, there exists a constant K such that:

$$\limsup_{N\to\infty} N \left|\mathbb{E}\left[h(X^N)\right] - h(x^*)\right| \leq K.$$

- Note: A similar result holds for the transient behavior.

Main ideas of the proof

1. Comparison of the generators:

$$(L^{(N)}h)(x) = \sum_{\ell \in \mathcal{L}} N\beta_\ell(x)(h(x+\frac{\ell}{N}) - h(x))$$
$$(\Lambda h)(x) = \sum_{\ell \in \mathcal{L}} \beta_\ell(x) Dh(x) \cdot \ell = Dh(x) \cdot f(x)$$

2. Stein's method :

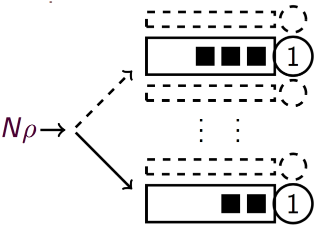
$$\mathbb{E}\left[h(X^N) - h(x^*)\right]\mathbb{E}\left[(\Lambda - L^{(N)})(Gh)X^N\right],$$

where $Gh(x) = \int_0^\infty (h(\Phi_t x) - h(x^*))dt$ satisfies $(x) - h(x^*) = \Lambda(Gh)x$.

3. Perturbation theory: $D^2(Gh)$ is twice-differentiable.

Outline

The two choice model[2]



Randomly choose two, and select one

Infinite state-space:

$$X_0(t), X_1(t), \dots$$

where

$X_i(t)$ = fraction with i or more jobs.

[2] This model or variants have been heavily studied (Vvedenskaya 96, Mitzenmacher 98 ... Tsitsiklis et al. 2016,...).

Does this model satisfies our assumptions?

$(\mathbf{E}, \|\cdot\|)$ is the set of infinite sequences such that $\|x\|_w = \sum_{i=1}^{\infty} w_i |x_i| < \infty$.

- Transitions : easy
- Regularity of the drift : easy
- Unique attractor : mitzenmacher 98
- Stationary measure concentrates on a bounded set : coupling argument : 2-choice $\ll$ 1-choice.

The power of two-choice

Our theory guarantees that the average queue length satisfies:

$$m^N(\rho) = m^\infty(\rho) + O(1/N),$$

where $m^\infty(\rho) \approx \log_2 \frac{1}{1-\rho}$ (as $\rho \to 1$).

The power of two-choice

Our theory guarantees that the average queue length satisfies:

$$m^N(\rho) = m^\infty(\rho) + O(1/N),$$

where $m^\infty(\rho) \approx \log_2 \frac{1}{1-\rho}$ (as $\rho \to 1$).

By simulation, we observe that $N(m^N(\rho) - m^\infty) = d(\rho) \approx \dfrac{\rho^2}{2(1-\rho)}$

N	10	20	30	50	$+\infty$
$m^N(\rho)$	2.804	2.566	2.491	2.434	–
$m^\infty(\rho) + \dfrac{\rho^2}{2N(1-\rho)}$	2.758	2.555	2.488	2.434	2.353

Table: Average queue length in the two-choice model ($\rho = 0.9$).

The quality of the approximation degrades as ρ goes to 1

Simulation results suggest that:

$$m^N(\rho) \approx \underbrace{m^\infty(\rho)}_{\approx \log_2 \frac{1}{1-\rho}} + \frac{1}{N} \underbrace{d(\rho)}_{\approx \frac{\rho^2}{2(1-\rho)}} + O(\frac{1}{N^2})$$

Conjecture: the power of two-choice holds if $N = \Omega(\frac{1}{1-\rho})$

Outline

Recap

1. Convergence of mean-field model is $O(1/N)$.
   - Works for transient and steady-state
   - Works for infinite-dimensional state space.

2. Our approach is to focus on the expected values

CLT: $O(1/\sqrt{N})$

X = empirical measure

Object property $\mathbb{E}\left[X^N\right]$

x_i (mean-field approx)

Our result: the difference is $O(1/N)$

In practice

For many mean-field models :

$$\mathbb{E}\left[X^N\right] \approx x + \frac{C}{N},$$

- C can be computed for one N and then interpolated.

This provides a new light for the two-choice.

Does it always work?

- Works for the model of Kurtz
- Also works for the "Benaim-Le Boudec" by using uniformization

But: it requires the drift to be twice-differentiable.

- (see counter-example on the paper)

Extension and open questions

- Multistable equilibria.
- Can we go to the order $O(1/N^2)$? It is useful?
- I assumed twice differentiable (and it is needed).
  - Can we do something in between for the steady-state?
- Non-homogeneous population.
  - *e.g.*, caching

Paper (and slides) are reproducible:
`https://github.com/ngast/meanFieldAccuracy`

Thank you!

http://mescal.imag.fr/membres/nicolas.gast

nicolas.gast@inria.fr

Mean-field and decoupling

Benaïm, Le Boudec 08 — *A class of mean field interaction models for computer and communication systems*, M.Benaïm and J.Y. Le Boudec., Performance evaluation, 2008.

Le Boudec 10 — *The stationary behaviour of fluid limits of reversible processes is concentrated on stationary points.*, J.-Y. L. Boudec. , Arxiv:1009.5021, 2010

Darling Norris 08 — *R. W. R. Darling and J. R. Norris*, Differential equation approximations for Markov chains, Probability Surveys 2008

G. 16 — *Construction of Lyapunov functions via relative entropy with application to caching*, Gast, N., ACM MAMA 2016

G. 16 — *Expected Values Estimated via Mean-field approximation are $1/N$ accurate*, Gast, N., SIGMETRICS 2017

Budhiraja et al. 15 — *Limits of relative entropies associated with weakly interacting particle systems.*, A. S. Budhiraja, P. Dupuis, M. Fischer, and K. Ramanan. , Electronic journal of probability, 20, 2015.

References (continued)

Optimal control and mean-field games:

G.,Gaujal Le Boudec 12 — *Mean field for Markov decision processes: from discrete to continuous optimization*, N.Gast,B.Gaujal,J.Y.Le Boudec, IEEE TAC, 2012

G. Gaujal 12 — *Markov chains with discontinuous drifts have differential inclusion limits.*, Gast N. and Gaujal B., Performance Evaluation, 2012

Lasry Lions — *Mean field games*, J.-M. Lasry and P.-L. Lions, Japanese Journal of Mathematics, 2007.

Tembine at al 09 — *Mean field asymptotics of markov decision evolutionary games and teams*, H. Tembine, J.-Y. L. Boudec, R. El-Azouzi, and E. Altman., GameNets 00

Applications: caches

Don and Towsley — *An approximate analysis of the LRU and FIFO buffer replacement schemes*, A. Dan and D. Towsley., SIGMETRICS 1990

G. Van Houdt 15 — *Transient and Steady-state Regime of a Family of List-based Cache Replacement Algorithms.*, Gast, Van Houdt., ACM Sigmetrics 2015