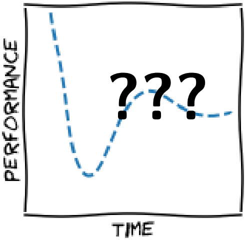
A *Refined* Mean Field Approximation

Nicolas Gast

Inria, Grenoble, France (joint work with Benny Van Houdt (Univ. Antwerp))

TU Eindhoven, February, 1st, 2018

This talk is about performance modeling of systems of interacting objects, using *stochastic* models

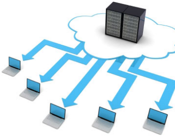


The main difficulty of probability : correlations

$$\mathbf{P}[A,B] \neq \mathbf{P}[A]\,\mathbf{P}[B]$$

Problem: state space explosion
S states per object, N objects $\Rightarrow S^N$ states

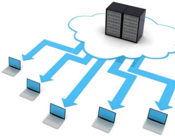
Can we study (Markovian) models of large systems?



$N = 1$: M/M/1!

N servers

Can we study (Markovian) models of large systems?



Model size grows like S^N :
intractable for $N \geq 10$

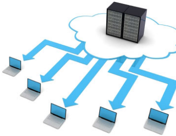
$N = 1$: M/M/1!

$N = 10$ $N = 100$ $N = 1000$

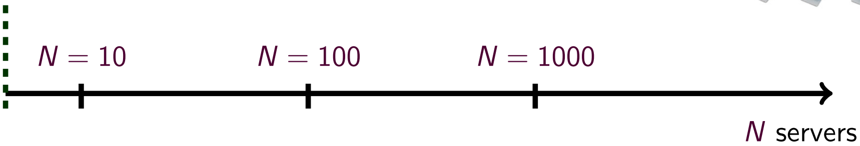
N servers

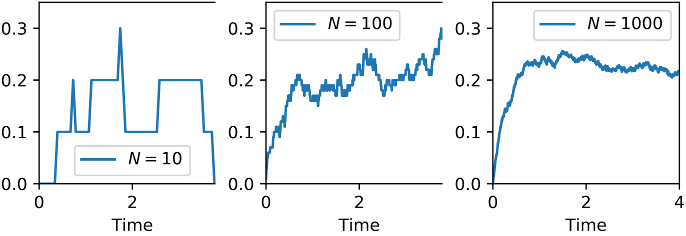
Can we study (Markovian) models of large systems?

Model size grows like S^N :
intractable for $N \geq 10$



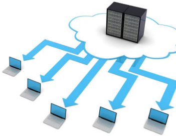
$N = 1$: M/M/1!



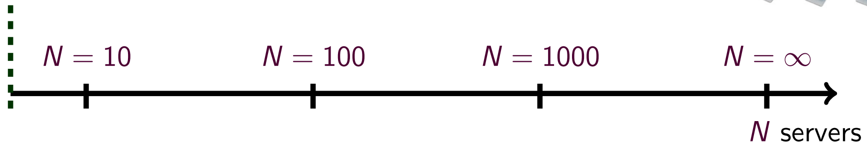


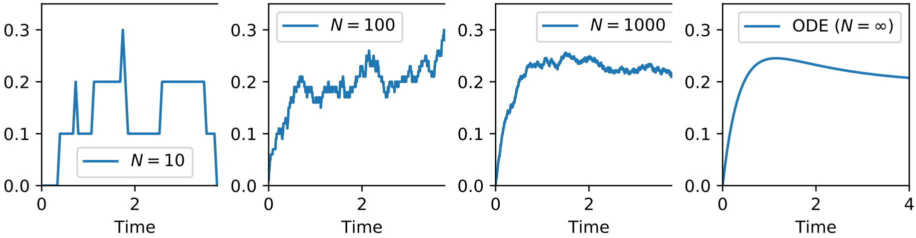
Can we study (Markovian) models of large systems?

Model size grows like S^N :
intractable for $N \geq 10$



$N = 1$: M/M/1!

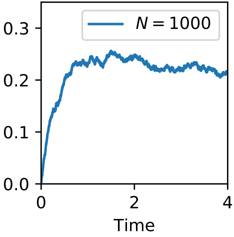




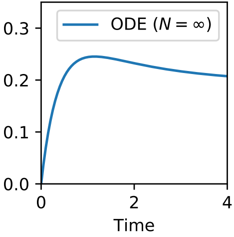
What happened is a law of large numbers

Some systems simplify as N goes to infinity : objects become independent

"Theorem". $\lim_{N\to\infty}$ (



) =

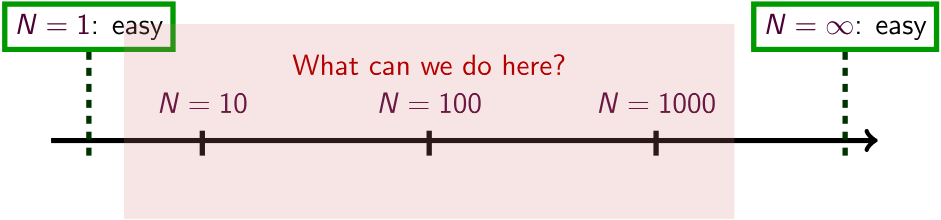


Mean field approximation

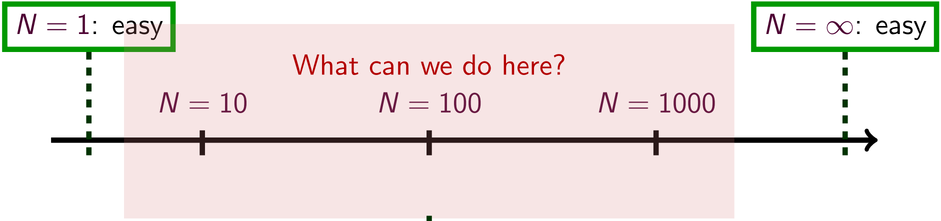
Mean field approximation has been shown to be asymptotically exact and has been successfully used in many context. For example :

- CSMA (see *e.g.* Thesis of F. Cecchi), 802.11 (Bianchi's formula)
- Load balancing (power of two-choice, Mitzenmacher 98 / Vvedenskaya 96, Tsitsiklis,Xu 2011& 2013)
- Caching algorithms (G and Van Houdt 2015)

We can study large systems. What about moderate sizes?



We can study large systems. What about moderate sizes?



purpose of this talk

For many systems, asymptotically:

$$Perf(N) \approx Perf(\infty) + \frac{1}{N}C$$

We can study large systems. What about moderate sizes?

$N = 1$: easy

What can we do here?

$N = 10$ $\quad$ $N = 100$ $\quad$ $N = 1000$

$N = \infty$: easy

purpose of this talk

For many systems, asymptotically:

$$Perf(N) \approx \boxed{Perf(\infty) + \frac{1}{N}C}$$

Mean field approximation

Refined mean field approximation

By studying what happens when $N \to \infty$, we get a very good approximation even for $N = 10$

	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

Outline

1. Kurtz' population model: classical convergence results
2. Accuracy of the approximation and refinement
3. In practice
4. Conclusion and recap

Outline

We study a population of N interchangeable objects.

X denotes the empirical measure.

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

CTMC

A continous-time Markov chain (CTMC) with state-space $\mathbf{E}$ 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 $\ell \in \mathcal{L}$):

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

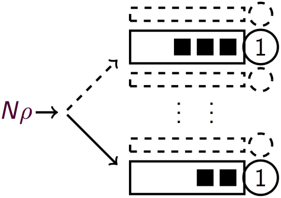
The drift is $f(x) = \sum_\ell \ell\beta_\ell(x)$.

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

Our running example will be the supermarket model

Vvedenskaya et al. 96, Mitzenmacher 98

Example: N servers

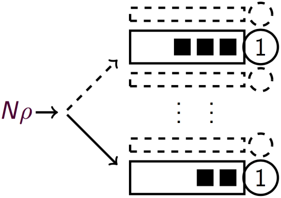


Randomly choose two, and select one

Our running example will be the supermarket model

Vvedenskaya et al. 96, Mitzenmacher 98

Example: N servers



Randomly choose two, and select one

X_i = fractions of servers with i or more jobs.

The transitions are:

$$X \mapsto X + \frac{1}{N}\mathbf{e}_i \text{ at rate } N\rho(X_{i-1}^2 - X_i^2)$$
$$X \mapsto X - \frac{1}{N}\mathbf{e}_i \text{ at rate } N(x_i - x_{i+1})$$

The mean field approximation is given by the (infinite) system of ODE:

$$\dot{x}_i = \underbrace{\rho(x_{i-1}^2 - x_i^2)}_{\text{arrivals}} - \underbrace{(x_i - x_{i+1})}_{\text{departures}}$$

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

Theorem (Kurtz 70s…Benaim-Le Boudec 08…Ying 16)

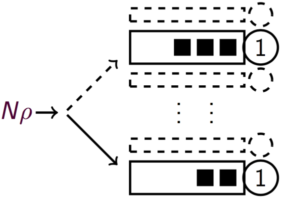
$$X^N(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$

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

Theorem (Kurtz 70s... Benaim-Le Boudec 08... Ying 16)

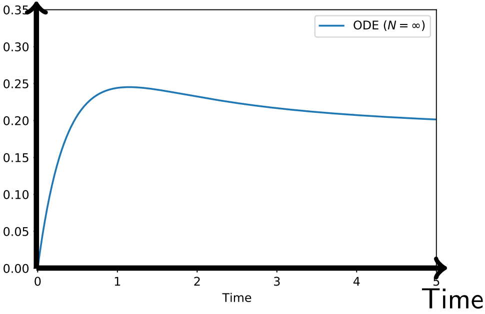
$$X^N(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$

Example: N servers



Randomly choose two, and select one

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

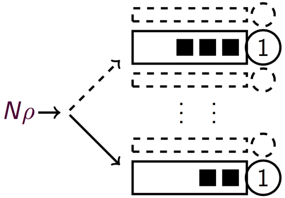


Idea of mean field: Some models simplify as $N \to \infty$

Theorem (Kurtz 70s... Benaim-Le Boudec 08... Ying 16)

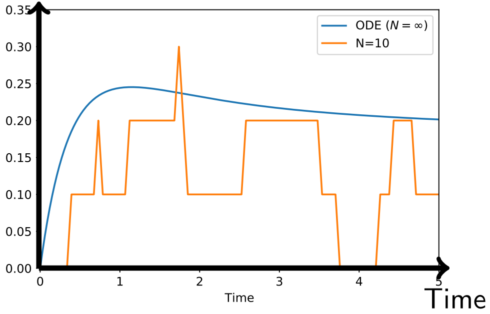
$$X^N(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$

Example: N servers



Randomly choose two, and select one

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

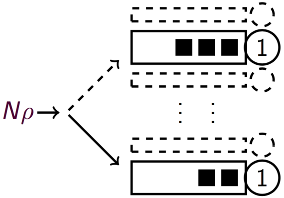


Idea of mean field: Some models simplify as $N \to \infty$

Theorem (Kurtz 70s…Benaim-Le Boudec 08…Ying 16)

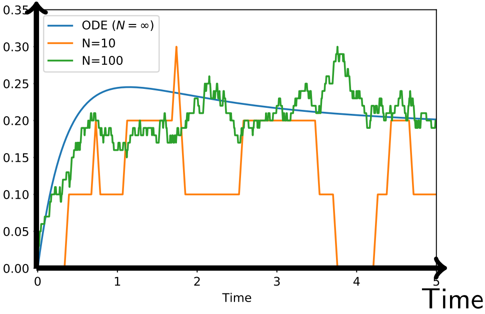
$$X^N(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$

Example: N servers



Randomly choose two, and select one

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

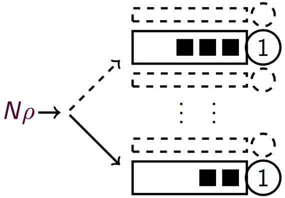


Idea of mean field: Some models simplify as $N \to \infty$

Theorem (Kurtz 70s… Benaim-Le Boudec 08… Ying 16)

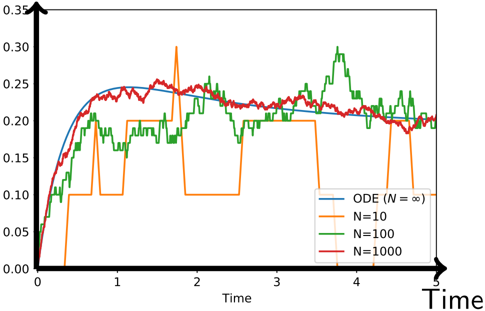
$$X^N(t) \approx x(t) + \frac{1}{\sqrt{N}} G_t$$

Example: N servers



Randomly choose two, and select one

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



In practice, can we use the approximation for $N = 10$? $N = 100$?

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, can we use the approximation for $N = 10$? $N = 100$?

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 $\rho = 0.9$

Under very general conditions:

1. The convergence is in $O(1/N)$, not $O(1/\sqrt{N})$

In practice, can we use the approximation for $N = 10$? $N = 100$?

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
Refined approximation	2.7513	2.3925	2.3566	2.3527

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

Under very general conditions:

1. The convergence is in $O(1/N)$, not $O(1/\sqrt{N})$
2. We can do better than mean field approximation.

Outline

Why is the accuracy of the approximation $1/N$ and not $1/\sqrt{N}$?

$$X_i = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{\{\text{Object } n \text{ is in state } i\}}$$

- Even if the objects were independent, the central limit theorem :

$$X_i = \underbrace{\mathbf{P}\left[\text{Object } n \text{ is in state } i\right]}_{\approx x_i \text{ (mean field approximation)}} + O\left(\frac{1}{\sqrt{N}}\right).$$

Why is the accuracy of the approximation $1/N$ and not $1/\sqrt{N}$?

$$X_i = \frac{1}{N}\sum_{n=1}^{N} \mathbf{1}_{\{\text{Object } n \text{ is in state } i\}}$$

- Even if the objects were independent, the central limit theorem :

$$X_i = \underbrace{\mathbf{P}\left[\text{Object } n \text{ is in state } i\right]}_{\approx x_i \text{ (mean field approximation)}} + O\left(\frac{1}{\sqrt{N}}\right).$$

- Our metric is different :

$$\begin{aligned}\mathbb{E}\left[X_i\right] &= \mathbf{P}\left[\text{Object } n \text{ is in state } i\right]. \\ &= x_i + \frac{C}{N} + O\left(\frac{1}{N^2}\right) \qquad \text{Our result}\end{aligned}$$

Steady-state analysis : main assumptions

(A0) $\sup_x \sum_\ell |\ell|^2 |\beta_\ell(x)| < \infty$.

(A1) The stochastic process is a density dependent population process.

(A2) The drift f is twice-differentiale

(A3) The ODE has a globally stable attractor π, *i.e.*, for any solution x of the ODE $\dot{x} = f(x)$:

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

(A4) For each N, the population process has a unique stationary distribution.

The constant can be easily evaluated numerically

The constant can be easily evaluated numerically

Let π be the fixed point of the mean field approximation and

$$A = Df(\pi) \qquad B = D^2 f(\pi) \qquad Q_{ij} = \sum_{\ell} \ell_i \ell_j \beta_\ell(\pi).$$

Let W be the unique solution of the Lyapunov equation

$$AW + (AW)^T = Q$$

THEOREM 3.1. *Assume that the model satisfies (A0–A4). Let $h : \mathcal{E} \to \mathbb{R}$ be a twice-differentiable function that has a uniformly continuous second derivative. Then,*

$$\lim_{N\to\infty} N \left(\mathbf{E}^{(N)}\left[h(X^{(N)})\right] - h(\pi)\right) = \sum_i \frac{\partial h}{\partial x_i}(\pi) V_i + \frac{1}{2}\sum_{i,j} \frac{\partial^2 h}{\partial x_i \partial x_j}(\pi) W_{ij}, \tag{2}$$

where the matrices A, C and W are defined above and V_i is equal to:

$$V_i = -\sum_j (A^{-1})_{i,j}\left[C_j + \frac{1}{2}\sum_{k_1,k_2} (B_j)_{k_1,k_2} W_{k_1,k_2}\right]. \tag{3}$$

Proof (1/2) – Comparison of generators

The generators of both systems are:

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

$$(\Lambda h)(x) = \sum_{\ell \in \mathcal{L}} \beta_\ell(x) Dh(x) \cdot \ell = Dh(x) \cdot f(x)$$

If h is C^2, then:

$$\lim_{N\to\infty} N(L^{(N)} - \Lambda)h(x) = \frac{1}{2} \sum_{\ell \in \mathcal{L}} \beta_\ell(x) D^2 h(x) \cdot (\ell, \ell)$$

Proof (2/2) – Stein's method (+ perturbation theory)

Let G_h be the function $G_h(x) = \int_0^\infty (h(\Phi_t(x)) - h(\pi))dt$, where $\Phi_t(x)$ is the solution of the ODE $\dot{x} = f(x)$ starting in x at time 0.
Then :

$$
\begin{aligned}
N\mathbb{E}\left[h(X^N) - h(\pi)\right] &= N\mathbb{E}\left[\Lambda G_h)(X^N)\right] \\
&= N\mathbb{E}\left[(\Lambda - L^{(N)})(G_h)(X^N)\right] \\
&= \frac{1}{2}\mathbb{E}\left[\sum_\ell \beta_\ell(X^N) D^2 G_h(X^N)\cdot(\ell,\ell)\right] + O(1/N) \\
&\to \frac{1}{2}\sum_\ell \beta_\ell(\pi) D^2 G_h(\pi)\cdot(\ell,\ell).
\end{aligned}
$$

The computation of $D^2 G_h(\pi)$ gives you the result (perturbation theory).

Outline

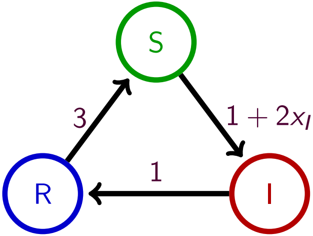
What can we do in practice?

$$Perf(N) \approx \underbrace{Perf(\infty) + \frac{C}{N}}_{\text{refined mean field approximation}}$$

- C cannot be computed in closed form very often.
- Numerical evaluation is easy, *e.g.*, `https://github.com/ngast/rmf_tool/`

	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

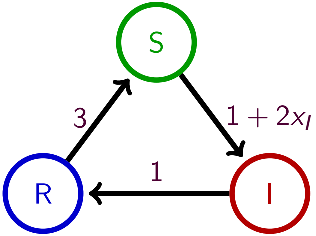
A simple example : SIR



The transitions are:

$$+\frac{1}{N}(-1,1,0) \text{ at rate } x_s + 2x_s x_I$$
$$+\frac{1}{N}(0,-1,1) \text{ at rate } x_I$$
$$+\frac{1}{N}(1,0,-1) \text{ at rate } 3x_R$$

A simple example : SIR



The transitions are:

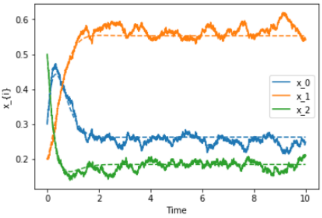
$$+\frac{1}{N}(-1,1,0) \text{ at rate } x_s + 2x_s x_I$$
$$+\frac{1}{N}(0,-1,1) \text{ at rate } x_I$$
$$+\frac{1}{N}(1,0,-1) \text{ at rate } 3x_R$$

In [1]:
```
# To load the library
import src.rmf_tool as rmf

# To plot the results
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:
```
# We create a "density dependent population process"
ddpp = rmf.DDPP()
# We then add the three transitions :
ddpp.add_transition([-1,1,0],lambda x:x[0]+2*x[0]*x[1])
ddpp.add_transition([0,-1,+1],lambda x:x[1])
ddpp.add_transition([1,0,-1],lambda x:3*x[2])
```

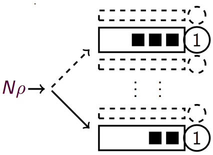
In [3]:
```
ddpp.set_initial_state([.3,.2,.5]) # initial state
ddpp.plot_ODE_vs_simulation(N=1000)
```



In [5]:
```
x = ddpp.fixed_point()
c = ddpp.theoretical_C()
print(x)
print(c)
```

```
[ 0.26259518  0.55305361  0.1843512 ]
[ 0.15875529 -0.11906646 -0.03968882]
```

More complex example : the two-choice model



Randomly choose two, and select one

The transitions are
(for $i \in \{1 \dots K\}$):

$$+\frac{1}{N}\mathbf{e}_i \text{ rate } N\rho(x_{i-1}^2 - x_i^2)$$
$$-\frac{1}{N}\mathbf{e}_i \text{ rate } N(x_i - x_{i+1})$$

```python
ddpp = rmf.DDPP()
K = 20 # we truncate at 20

# The vector 'e(i)' is a vector where only the $i$th coordinate equals $1$
def e(i):
    l = np.zeros(K)
    l[i] = 1
    return(l)

# We then add the transitions :
for i in range(K):
    if i>=1:
        ddpp.add_transition(e(i),
            eval('lambda x: rho*(x[{}]*x[{}] - x[{}]*x[{}] )'.format(i-1,i-1,i,i) ))
    if i<K-1:
        ddpp.add_transition(-e(i),
            eval('lambda x: (x[{}] - x[{}])'.format(i,i+1) ))
ddpp.add_transition(e(0), lambda x : rho*(1-x[0]*x[0]))
ddpp.add_transition(-e(K-1), lambda x : x[K-1])
```

```python
ddpp.set_initial_state(e(0)) # initial state

print('\t\t      N=10\t    N=50\t   N=inf',end='')
for rho in [0.7,0.9,0.95]:
    print('\nrho=',rho,'\t',end=' ')
    x = ddpp.fixed_point()
    c = ddpp.theoretical_C()
    for N in 10,50,np.inf:
        print(sum(x+c/N),end=' ')
```

```
                 N=10          N=50          N=inf
rho= 0.7       1.21502419299 1.14709894998 1.13011763922
rho= 0.9       2.75129433831 2.43238017933 2.35265163959
rho= 0.95      4.10172926564 3.39146081504 3.21389370239
```

For the two-choice (and many models), the quality of the approximation degrades as ρ approaches 1

The average queue length satisfies:

$$m^N(\rho) = \Theta_{\rho\to 1}\Big(\log \frac{1}{1-\rho}\Big) + O(1/N)$$

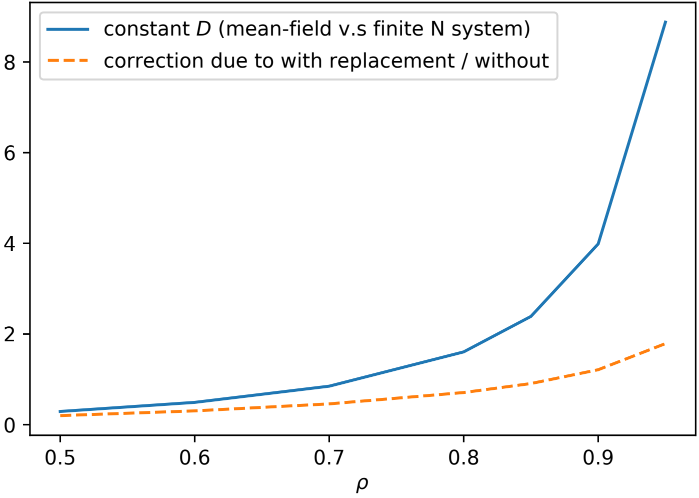
For the two-choice (and many models), the quality of the approximation degrades as ρ approaches 1

The average queue length satisfies:

$$m^N(\rho) = \Theta_{\rho\to 1}\Big(\log \frac{1}{1-\rho}\Big) + \frac{1}{N} \underbrace{\Theta_{\rho\to 1}\Big(\frac{1}{1-\rho}\Big)}_{\text{order of magnitude larger}} + O\Big(\frac{1}{N^2}\Big)$$

(based on a numerical evaluation of the $c(\rho) \approx \dfrac{\rho^2}{2}\dfrac{1}{1-\rho}$).

Power of two-choice : the impact of with/without replacement



Outline

Recap

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

2. We can use the rate of convergence to define a refined approximation. The main ideas are:
   - It is easy to compute $x = \lim_{N\to\infty} X^N$
   - It is easy to compute $C = \lim_{N\to\infty} N(X^N - \pi)$
   - The new approximation is $x + C/N$.

The refined approximation is often accurate even for $N = 10$:

	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

To go further :

More examples in the paper, *e.g.*: two-choice with/without replacement.

Open questions

- Multistable equilibria.
- Can we go to the order $O(1/N^2)$? It is useful?
- Non-homogeneous population, *e.g.*, caching

Main references :

- A Refined Mean Field Approximation by Gast and Van Houdt. To appear in SIGMETRICS 2018 https://hal.inria.fr/hal-01622054/ https://github.com/ngast/rmf_tool/
- Expected Values Estimated via Mean Field Approximation are O(1/N)-accurate by Gast. SIGMETRICS 2017. 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	*mean field approximation is $1/N$ accurate*, Gast, N., submitted
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