

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

Nicolas Gast

▶ **To cite this version:**

Nicolas Gast. Expected Values Estimated via Mean-Field Approximation are 1/N-Accurate. ACM SIGMETRICS / International Conference on Measurement and Modeling of Computer Systems SIGMETRICS '17 , Jun 2017, Urbana-Champaign, United States. pp.26, 10.1145/2896377.2901463 . hal-01553133

HAL Id: hal-01553133

https://inria.hal.science/hal-01553133

Submitted on 4 Jul 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

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

NICOLAS GAST, Inria

Mean-field approximation is a powerful tool to study large-scale stochastic systems such as data-centers – one example being the famous power of two-choice paradigm. It is shown in the literature that under quite general conditions, the empirical measure of a system of N interacting objects converges at rate $O(1/\sqrt{N})$ to a deterministic dynamical system, called its mean-field approximation.

In this paper, we revisit the accuracy of mean-field approximation by focusing on expected values. We show that, under almost the same general conditions, the expectation of any performance functional converges at rate $O(1/N)$ to its mean-field approximation. Our result applies for finite and infinite-dimensional mean-field models. We also develop a new perturbation theory argument that shows that the result holds for the stationary regime if the dynamical system is asymptotically exponentially stable. We provide numerical experiments that demonstrate that this rate of convergence is tight and that illustrate the necessity of our conditions. As an example, we apply our result to the classical two-choice model. By combining our theory with numerical experiments, we claim that, as the load ρ goes to 1, the average queue length of a two-choice system with N servers is $\log_2 \frac{1}{1-\rho} + \frac{1}{2N(1-\rho)} + O\left(\frac{1}{N^2}\right)$.

CCS Concepts: • **General and reference** → **Performance**; • **Mathematics of computing** → **Stochastic processes**; *Queueing theory*; Ordinary differential equations;

Additional Key Words and Phrases: Mean-field approximation; queueing theory; accuracy of approximation; supermarket model; power-of-two-choice

ACM Reference format:
Nicolas Gast. 2017. Expected Values Estimated via Mean-Field Approximation are 1/N-Accurate. *Proc. ACM Meas. Anal. Comput. Syst.* 1, 1, Article 16 (June 2017), 26 pages.
https://doi.org/http://dx.doi.org/10.1145/3084454

1 INTRODUCTION

Mean-field approximation is a powerful tool for studying systems composed of a large number of interacting objects. The idea of mean-field approximation is to replace a complex stochastic system by a simpler deterministic dynamical system. This dynamical system is constructed by considering that each object interacts with an average of the other objects (the *mean*-field). When each object has a finite or countable state-space, this dynamical system is usually a non-linear ordinary differential equation (ODE) $\dot{x} = f(x)$, where $x_i(t)$ denotes the fraction of objects in state i at time t and f is the drift of the system.

Mean-field approximation is widely used to study the performance of computer-based systems: queueing networks [1], wireless networks [5], dissemination algorithms [6], caching [12], SSDs [24],... An important area of application is the analysis of resource allocation strategies in server farms or data centers: such a

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2017 Copyright held by the owner/author(s). Publication rights licensed to Association for Computing Machinery.
2476-1249/2017/6-ART16 $15.00
https://doi.org/http://dx.doi.org/10.1145/3084454

system is composed of a large number of servers that interact because of scheduling or routing strategies [10, 17, 19, 20, 22, 25]. A typical example is the *power of two-choices*: Mean-field approximation has been used in [20, 25] to show that routing a task to the least loaded of two randomly sampled servers significantly reduces the response time compared to a purely random allocation.

Mean-field approximation is known to be asymptotically exact for many systems. For these systems, the fraction of objects in a given state i at time t, $X_i^{(N)}(t)$, converges to a deterministic quantity $x_i(t)$, as the number of objects N goes to infinity. The rate of convergence of $X_i^{(N)}$ to x_i has been studied by several papers, *e.g.* [2, 15, 26], that show that the expected distance between $X^{(N)}$ and x is of the order of $1/\sqrt{N}$:

$$\mathbb{E}\left[\left\|X^{(N)} - x\right\|\right] \approx \frac{C}{\sqrt{N}}. \tag{1}$$

This result is a like a central-limit-theorem for mean-field systems: $X^{(N)}(t)$ is equal to $x(t)$ plus $1/\sqrt{N}$ times a Gaussian noise [15]. It was originally proven for finite time-horizon and recently extended to stationary distributions in [26].

Yet, we believe that Equation (1) does not fully explain the accuracy of mean-field approximation. As an example, we provide in Table 1 results obtained by simulation on how the mean-field approximation is accurate for the power of two-choices model[1] of [20, 25]. We report the average queue length in steady-state as a function of the number of servers N for $\rho = 0.9$. We observe that the error made by the mean-field approximation decreases as $1/N$, much faster than the $1/\sqrt{N}$ suggested by Equation (1).

Number of servers (N)	10	100	1000	$+\infty$
Average queue length (m^N)	2.81	2.39	2.36	2.35
Error ($m^N - m^\infty$)	0.45	0.039	0.004	0

Table 1. Average queue length in the two-choice model. The values for a finite number of servers N are obtained by simulation. The value for $N = +\infty$ is the mean-field approximation.

This discrepancy comes from the fact that the error term $m^N - m^\infty$ is the distance between the mean-field value m^∞ and an expected function of $X^{(N)}$. It is not the expectation of a distance as in Equation (1). We therefore adopt a new point of view in this paper. Instead of studying the distance between $X^{(N)}(t)$ and $x(t)$ as in Equation (1), we study the distance between the expectation of a function of $X^{(N)}(t)$ and its mean-field approximation. As a norm is a convex function, we expect the former to be smaller than the latter. We will show in fact that there is an order of magnitude of difference: under mild conditions, for any function h that is twice differentiable, this distance is of order $1/N$:

$$\left|\mathbb{E}\left[h(X^{(N)})\right] - h(x)\right| = O\left(\frac{1}{N}\right). \tag{2}$$

As for Equation (1), we show that Equation (2) holds for the transient regime and can be extended to the stationary regime under the same conditions as [26]. As a byproduct, Equation (1) can be recovered from Equation (2) by using $h(.) = \|. - x\|^2$.

This result shows that an average value estimated via mean-field approximation is $1/N$-accurate. In a queuing network such as the two-choice model, the average queue length can be expressed as $\mathbb{E}[h(X^{(N)})]$. Equation (2) shows that the average queue length converges at rate $O(1/N)$ to its mean-field approximation. This is what is observed in Table 1, where $\mathbb{E}[h(X^{(N)})] \approx h(x) + 4/N$.

[1] We define and study this model in more details in Section 5.

Contributions. We prove our result for a generalization of the classical population model of Kurtz in which each object can have a countable state-space.

The classical methods to obtain Equation (1) for transient [2, 15] or stationary regime [3] rely on martingale concentration arguments. Our approach is different and uses an idea inspired by Stein's method and [14, 21, 26] to relate Equation (2) and the convergence of the generators of the Markov chain.

From a theoretical perspective, the main contribution of our paper is to provide a unified framework in which Equation (2) is true: our results hold for the transient and stationary regime and for finite or infinite-dimensional models. Using Stein's method, we show that the $O(1/N)$ rate holds if the derivative with respect to its initial condition of $\int_0^t \Phi_s x ds$ exists and is Lipschitz-continuous (where $t \mapsto \Phi_t x$ denote the unique solution of the ODE $\dot{x} = f(x)$ that starts in x at time 0). We show that, for the transient regime, this holds as soon as the drift f has a derivative that is Lipschitz-continuous. For the stationary regime, it holds when, in addition, the ODE has a unique attractor that is exponentially stable. In the transient case, [14] obtains a $O(1/N)$ like Equation (2) for a simplified version of our model where a transition affects at most one object. For the stationary regime, a $O(1/\sqrt{N})$ like Equation (1) is obtained in [26]. Our proof is inspired by the methodology of [26] but we obtain a stronger result by working with a generic function h. Note that infinite-dimensional models arise naturally when one considers queuing systems with unbounded queues. They are not considered in [14, 26].

Another contribution of our paper compared to [14, 26] is to characterize what happens when the derivative of the drift is not Lipschitz-continuous. We show that if the drift is differentiable and its derivative is α-Hölder continuous, then the convergence rate occurs at rate $O(1/\sqrt{N}^{1+\alpha})$. We provide a numerical example in Section 4 that shows that in general this exponent is tight and that having a Lipschitz-continuous drift is not sufficient to obtain Equation (2). This contrasts with the convergence rate of Equation (1) that is known to be $O(1/\sqrt{N})$ as soon as the drift is Lipschitz-continuous (or one-sided Lipschitz-continuous) for the transient case. Hence, only Lipschitz-continuity (or one-sided Lipschitz-continuity as in [11, 22]) is not sufficient to guarantee a $1/N$-accuracy of the mean-field estimates.

From a performance analysis perspective, one of the main insights provided by our result is that for many mean-field models and performance functions h, we have

$$\mathbb{E}\left[h(X^{(N)})\right] = h(x) + \frac{C}{N} + o\left(\frac{1}{N}\right). \tag{3}$$

This provides a refinement of the mean-field approximation: the performance of a system of size N can be precisely estimated from two quantities: its mean-field approximation $h(x)$ and the constant C corresponding to the error of the mean-field approximation. The constant C can be estimated analytically or by simulation for a small system size and then extrapolated to larger system sizes. As an example, we study in detail the convergence rate of the classical power-of-two-choice model of [20, 25]. This model shows that routing an incoming job to the shortest of two randomly sampled queues greatly reduces the average queue length : as N goes to infinity, for a system with load $\rho \approx 1$, the queue length equals $\Theta(\log 1/(1-\rho))$ compared to $1/(1-\rho)$ in the case of purely random allocation. We show that for any fixed ρ, the average queue length in a N-server two-choice system is at distance d_ρ/N of its mean-field approximation. We provide numerical evidence that this d_ρ grows as $1/(1-\rho)$ as ρ goes to one. As a result, we conjecture that the average queue length in a N-server two-choice system is

$$m^N(\rho) = \Theta_{\rho\to 1}\left(\log \frac{1}{1-\rho}\right) + \frac{1}{N}\Theta_{\rho\to 1}\left(\frac{1}{1-\rho}\right) + O\left(\frac{1}{N^2}\right),$$

where $\Theta(1/(1-\rho))$ is equivalent to $1/(2-2\rho)$ as ρ goes to 1.

The above equation is constructed by assuming first that N goes to infinity and then that ρ goes to one. This result should be contrasted with the heavy-traffic limit of [9], who study the case where the load depends on N and goes to one as N goes to infinity: $\rho_N = 1 - 1/\eta_N$ with $\lim_{N\to\infty} \eta_N = \infty$. The authors show that the mean-field

approximation always under-estimates the average queue length. They also establish a heavy-traffic regime when η_N grows sub-linearly. Our result suggests that for this scaling, the difference between the average queue length of the system of size N and its mean-field limit is of order $1/(N(1-1\rho_N)) = \eta_N/N \to 0$. A comparison of both results is not straightforward as the considered metric is different. Yet, our result confirms that when η_N grows sub-linearly, there is little difference between the system of finite N and the mean-field.

Reproducibility. The code to reproduce the paper – including simulations, figures and text – is available at https://github.com/ngast/meanFieldAccuracy.

Roadmap. The paper is organized as follows. We introduce the model in Section 2. We provide the main theoretical results in Section 3: the transient regime and its extension to stationary regime when the ODE has a unique exponentially stable attractor. We illustrate the necessity of the regularity assumptions in Section 4. We apply these results to the two-choice model in Section 5. Section 6 contains the proof of the most technical lemmas. Finally, we conclude in Section 7.

2 MEAN-FIELD MODEL

2.1 Population Processes

We consider population processes described by the classical model of density-dependent population process of [15, 16]. We recall the definition here. A population process is a sequence of continuous-time Markov chains (CTMC) $(X^{(N)})$. For each N, $X^{(N)}$ evolves on a subset of some Banach space[2] $(\mathcal{E}, \|.\|)$. We assume that there exists a set of vectors $\mathcal{L} \in \mathcal{E}$ and a set of functions $\beta_\ell : \mathcal{E} \to \mathbb{R}^+$ such that $X^{(N)}$ jumps from x to $x + \ell/N$ at rate $N\beta_\ell(x)$ for each $\ell \in \mathcal{L}$.

We will refer to the chain $X^{(N)}$ as the system of size N, although N does not *a priori* correspond to the system's size. For such a system, we define the drift f as

$$f(x) = \sum_{\ell \in \mathcal{L}} \ell \beta_\ell(x)$$

This definition means that $f(x)dt$ is the expected variation of a chain $X^{(N)}$ that would start in state x during a small time interval dt. It does not depend on N.

Mean-field interacting system. A particularly interesting subclass of population processes is the case of mean-field interacting systems. Such a system is composed of N objects. Each object lives in a finite or countable state-space $\mathcal{S}$. The state of the object n at time t is denoted by $S_n^{(N)}(t)$. Let $X^{(N)}(t)$ be the empirical measure at time t: for $i \in \mathcal{S}$, $X_i^{(N)}(t)$ is the fraction of objects in state i at time t:

$$X_i^{(N)}(t) := \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{\{S_n^{(N)}(t)=i\}}.$$

We say that this system is a mean-field interacting system if $(X^{(N)})$ is a population process. The process $X^{(N)}$ lives in $\mathcal{P}(\mathcal{S})$, the set of probability measures on $\mathcal{S}$.

Note that if all objects are exchangeable, we have

$$\begin{aligned} \mathbb{E}\left[X_i^{(N)}(t)\right] &= \frac{1}{N} \sum_{n=1}^{N} \mathbf{P}\left[S_n^{(N)}(t) = i\right] \\ &= \mathbf{P}\left[S_1^{(N)}(t) = i\right]. \end{aligned} \tag{4}$$

[2]Typically, we will use $\mathcal{E} = \mathbb{R}^d$ or $\mathcal{E} = \mathbb{R}^{\mathbb{N}}$ (the set of infinite sequences) with an appropriate norm.

The two-choice model is an example of mean-field interacting system (see Section 5).

Imperfect population processes. In some cases, it is useful to consider models in which the set of transitions $\mathcal{L}$ or the transition functions β_ℓ vary with the system size. We refer to this case as the *imperfect* population processes and we add a superscript N to all quantities that depend on N. The drift of an imperfect model depends on N and will be denoted by f^N, with $f^N(x) = \sum_{\ell \in \mathcal{L}^N} \ell \beta_\ell^N(x)$.

2.2 Notations

For a function $h : \mathcal{E} \to \mathcal{F}$, we will denote by Dh the derivative of h (if it exists). For $\alpha \in (0, 1]$, a function h is α-Hölder continuous with constant L if for all $x, y \in \mathcal{E}$:

$$\|h(x) - h(y)\| \leq L\|x - y\|^\alpha.$$

Lipschitz-continuity corresponds to the case $\alpha = 1$: a function is Lipschitz-continuous if it is 1-Hölder continuous.

The set $C^{1+\alpha}(\mathcal{E}, \mathcal{F})$ denotes the set of functions $h : \mathcal{E} \to \mathcal{F}$ that are differentiable and such that the first derivative of h is α-Hölder continuous. The set $C^1(\mathcal{E}, \mathcal{F})$ denotes the set of functions $h : \mathcal{E} \to \mathcal{F}$ that are differentiable and such that the first derivative of h is continuous and bounded.

In the theorems, we assume that f is Lipschitz-continuous. This implies that, for any initial condition the ODE $\dot{x} = f(x)$ has a unique solution. For $t \geq 0$, we denote by $\Phi_t x$ the value at time t of the solution that starts in x at time 0. It satisfies

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

We denote by $L^{(N)}$ the generator of the system of size N and by Λ the generator of the ODE. They associate to each function h that is differentiable, two functions $L^{(N)}h$ and Λh that are defined as

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

By definition of the drift: $(\Lambda h)(x) = \sum_{\ell \in \mathcal{L}} \beta_\ell(x) Dh(x) \cdot \ell$.

3 CONVERGENCE RESULTS

The classical approach to show the convergence of a stochastic system to its mean-field approximation is to obtain a bound on the expected distance between $X^{(N)}(t)$ and its deterministic limit $\Phi_t x$. Our approach is different in the sense that we obtain a bound of the distance between the expectation of $X^{(N)}(t)$ and $\Phi_t x$. We show that the latter is in general much smaller than the former:

$$O\left(\frac{1}{N}\right) = \left\|\mathbb{E}\left[X^{(N)}(t)\right] - \Phi_t x\right\|$$
$$\ll \mathbb{E}\left[\left\|X^{(N)}(t) - \Phi_t x\right\|\right] = O\left(\frac{1}{\sqrt{N}}\right).$$

3.1 Transient Regime

Our first result concerns the transient regime. We show that in this case, the Lipschitz-continuity of the derivative of the drift suffices to show that an average value estimated by the mean-field approximation – $h(\Phi_t x)$ – is at distance $O(1/N)$ from the true value – $\mathbb{E}[h(X^{(N)}(t))]$. One application of this result is when one wants to

compute the marginal law of an object of a mean-field interaction model. This result combined with Equation (4) implies that the marginal law of one object converges at rate $O(1/N)$ to its mean-field approximation:

$$\left|\mathbb{P}\left(S_1^{(N)}(t)=i\right)-x_i(t)\right| \leq \frac{C(t)}{N},$$

where $x(t)$ denotes the value at time t of the unique solution of the ordinary differential equation starting in $x \approx X^{(N)}(0)$. This result complements functional central limit theorems – such as the results of [16] – that show that the sample paths of $X^{(N)}$ are at distance $O(1/\sqrt{N})$ of x.

Note that the conditions to apply Theorem 3.1 are easy to verify and can be done by a syntactic analysis of the model. In Section 3.4, we show that this result can be modified in the case of drifts whose derivatives are α-Hölder continuous.

Theorem 3.1. *Let $X^{(N)}$ be a population process on a Banach space $\mathcal{E}$ such that $\sup_x \sum_{\ell\in\mathcal{L}} \beta_\ell(x)\|l\|^2 < \infty$. Assume that the derivative of the drift exists and that both the drift and its derivative are Lipschitz-continuous. Let $h:\mathcal{E}\to\mathbb{R}$ be a differentiable function such that both h and its derivative are Lipschitz-continuous with constant L. Assume that $\|\mathbb{E}[X^{(N)}(0)]-x\| \leq c/N$. Then, for each t, there exists a constant $C(t)<\infty$ such that*

$$\left|\mathbb{E}\left[h(X^{(N)}(t))\right]-h(\Phi_t x)\right| \leq \frac{C(t)L}{N}. \tag{5}$$

Proof. The proof can be decomposed in three steps. The first step is to reduce Equation (5) to a convergence of the generators. For that, we use Lemma 6.1 that shows

$$\mathbb{E}\left[\left|h(X^{(N)}(t))-h(\Phi_t x)\right|\right] \leq \int_0^t \sup_{x\in\mathcal{E}} |(\Lambda - L^{(N)})h(\Phi_s x)|ds. \tag{6}$$

The second step is to remark that if the derivative of a function g is Lipschitz-continuous with constant L, then $L^{(N)}g$ converges to Λg at rate N^{-1}. Indeed:

$$\begin{aligned}\left\|(L^{(N)}-\Lambda)gx\right\| &\leq \sum_{\ell\in\mathcal{L}}\beta_\ell(x)\left\|N(g(x+\frac{\ell}{N})-g(x))-Dg(x)\cdot\ell\right\| \\ &\leq \frac{L}{2}\sum_{\ell\in\mathcal{L}}\beta_\ell(x)\frac{\|\ell\|^2}{N}.\end{aligned} \tag{7}$$

The third step is provided by Lemma 6.2 which shows that when the derivative of the drift is Lipschitz-continuous, the derivative of the function $x\mapsto\Phi_s x$ exists and is also Lipschitz-continuous with constant A_s. Hence, the derivative of $x\mapsto h(\Phi_s x)$ is Lipschitz-continuous with some constant LA_s. We can then plug Equation (7) into (6) to show

$$\mathbb{E}\left[\left|h(X^{(N)}(t))-h(\Phi_t x)\right|\right] \leq \frac{L\sup_{s\leq t}A_s}{2LN}\sup_{x\in\mathcal{E}}\sum_{\ell\in\mathcal{L}}\beta_\ell(x)\|\ell\|^2.$$

The statements and the proofs of the two lemmas are postponed to Section 6.1. □

3.2 Stationary Regime

We now focus on the stationary distribution of $X^{(N)}$ and how it concentrates on the stationary regime of the ODE.

It is shown in [2] that if the ODE has a unique attractor (*i.e.* a fixed point to which all trajectories converge), then the stationary measure of the system of size N concentrates on this fixed point x^* as N goes to infinity. In this section, we show that the rate at which it concentrates is $O(1/N)$ under the condition that the ODE has a unique attractor that is locally exponentially stable (as in [26]).

We say that a point x^* is a locally exponentially stable attractor if there exists a bounded neighborhood V of x^*, and two constants $a, b > 0$ such that

(A1) For $x \in V$: $\|\Phi_t x - x^*\| \leq be^{-at}\|x - x^*\|$ (x^* is locally exponentially stable).
(A2) For all $x \in \mathcal{E}$, there exists T such that for $t \geq T$: $\Phi_t(x) \in V$ (the neighborhood V attracts all trajectories).

In practice, condition (A1) is easy to verify, as it depends mainly on whether the ODE is locally stable. This can be done by a numerical procedure that linearizes the drift f around its fixed point and computes the eigenvalues of its Jacobian matrix. The most difficult condition to verify is (A2): proving that the fixed point of an ODE is a global attractor is difficult because it depends not only on the form of the ODE but also on precise values of its parameters [2, 7]. Note (A2) is necessary to apply the classical concentration results of [2] which means that our conditions are not more difficult to verify than the ones of [2].

We assume that for N large enough, the stationary measure of the system of size N exists. Moreover, in order to deal with possibly non-compact state-space $\mathcal{E}$, we assume that, as N goes to infinity, the stationary measure concentrates on a bounded set $\mathcal{B}$ of $\mathcal{E}$ at rate $O(1/N^2)$, *i.e.*

(A3) There exists a bounded set $\mathcal{B}$ and a constant $B > 0$ such that $\mathbf{P}[X^{(N)} \notin \mathcal{B}] \leq B/N^2$ (in steady-state).

If the system evolves in a compact space, assumption (A3) is trivially satisfied by taking $\mathcal{B} = \mathcal{E}$. In the case of unbounded state-space $\mathcal{E}$, we know from [2] that the stationary measure converges on the Dirac measure in x^* as N goes to infinity. The condition (A3) is a bound on how far $X^{(N)}$ can deviate from x^*. For the two-choice model, we show it by bounding $X^{(N)}$ by a system of N independent $M/M/1$ queues and then by using a large-deviation type results (see the end of Section 6.4).

Theorem 3.2. *In addition to the assumptions of Theorem 3.1, assume that*

- *the ODE has a unique attractor x^* that is locally exponentially stable (assumptions (A1) and (A2))*
- *either $\mathcal{E}$ is bounded or the stationary measure concentrates on a bounded set $\mathcal{B}$ at rate $O(1/N^2)$ (Assumption (A3)).*

Then for any $h \in C^2(\mathcal{E}, R)$ bounded, the constant $C(t)$ in Theorem 3.1 is uniformly bounded (in t). In particular, there exists a constant $C < \infty$, such that:

$$\left|\mathbf{E}^{(N)}\left[h(X^{(N)})\right] - h(x^*)\right| \leq \frac{C}{N} \tag{8}$$

where $\mathbf{E}^{(N)}[.]$ denotes the expectation with respect to the stationary distribution of the system of size N.

Proof. Without loss of generality, we assume that $h(x) = h(x^*)$ outside a bounded set. Indeed, if this is not the case, we may replace h by a function h' that is equal to h on $\mathcal{B}$ and equal to $h(x^*)$ outside a bounded set. In which case, by Assumption (A3) and because h is bounded, $\mathbf{E}^{(N)}[h(X^{(N)})] = \mathbf{E}^{(N)}[h'(X^{(N)})] + O(1/N^2)$.

The first step of the proof is then to construct Equation (10). This equation is the analog of (6) as $t \to \infty$. Following Stein's method and [26], we define the function Gh as

$$Gh : x \mapsto \int_0^\infty (h(\Phi_t x) - h(x^*))\, dt.$$

The previous integral is well defined because x^* is exponentially stable. By differentiating in $s = 0$ the quantity $\int_s^\infty (h(\Phi_t x) - h(x^*))\, dt$ with respect to s, it can be shown that Gh satisfies the Poisson equation:

$$h(x^*) - h(x) = \Lambda Gh(x) \tag{9}$$

As $L^{(N)}$ is the generator associated to the stationary measure, for any function h' such that $\mathbf{E}^{(N)}[\|h'(X^{(N)})\|]$ exists, we have $\mathbf{E}^{(N)}[L^{(N)}h'(X^{(N)})] = 0$. Applying this to the function $h'(x) = Gh(x)$, and combining this with

(9), we get:

$$\begin{aligned}\mathbf{E}^{(N)}\left[h(x^*) - h(x)\right] &= \mathbf{E}^{(N)}\left[\Lambda G h(X^{(N)})\right] \\ &= \mathbf{E}^{(N)}\left[(\Lambda - L^{(N)})Gh(X^{(N)})\right] \end{aligned} \tag{10}$$

The previous equation is essentially a generalization of [26, Equation (4) and Equation (7)].

We then decompose Equation (10) in two terms:

$$\begin{aligned}\mathbf{E}^{(N)}\left[h(x^*) - h(x)\right] = \mathbf{E}^{(N)}&\left[(\Lambda - L^{(N)})Gh(X^{(N)})\mathbf{1}_{X^{(N)}\in\mathcal{B}}\right] \\ &+ \mathbf{E}^{(N)}\left[(\Lambda - L^{(N)})Gh(X^{(N)})\mathbf{1}_{X^{(N)}\notin\mathcal{B}}\right]\end{aligned}$$

As shown by Equation (7), the first term of this equation converges at rate $O(1/N)$ to 0 if the derivative of Gh exists and is Lipschitz-continuous on $\mathcal{B}$. This is what we show in Lemma 6.3. The second term can be bounded by using (A3), as shown in Lemma 6.4. The proofs of both lemmas are postponed to Section 6.2. □

3.3 Extension to Imperfect Models

The above results apply for exact mean-field models. For some mean-field models, the set of transitions $\mathcal{L}^N$ or the rates β_ℓ depend on the system size. In this section, we show that the above results can be extended to the case of imperfect models. Note that this theorem does not require that the convergence of the set of transitions of the rate functions but only that f^N, the drift of the system of size N, converges uniformly to a function f.

Theorem 3.3. *Assume that for each N, the function f^N is differentiable and has a Lipschitz-continuous derivative. Assume that both f^N and its derivative Df^N converge uniformly to f and Df.*

(i) *If the other assumptions of Theorem 3.1 are satisfied, then Theorem 3.1 holds, replacing Equation (5) by*

$$\mathbb{E}\left[h(X^{(N)}(t)) - h(\Phi_t X^{(N)}(0))\right] \leq C(t)L\left(\frac{1}{N} + \left\|f^N - f\right\|\right).$$

(ii) *If the other assumptions of Theorem 3.2 are satisfied, then Theorem 3.2 holds, replacing Equation (8) by*

$$\mathbf{E}^{(N)}\left[h(X^{(N)}) - h(x^*)\right] \leq CL\left(\frac{1}{N} + \left\|f^N - f\right\|\right).$$

Proof. One key ingredient of the previous Theorems is Equation (7), that shows that $L^{(N)}$ converges to Λ at rate $O(1/N)$. In an imperfect population process, the generator of the system of size N is

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

Let $\Lambda^{(N)}$ be the generator of the ODE associated to f^N:

$$\Lambda^{(N)}h(x) = Dh(x)\cdot f^N(x)$$

It follows that

$$\left\|(L^{(N)} - \Lambda)hx\right\| \leq \left\|(L^{(N)} - \Lambda^{(N)})hx\right\| + \left\|(\Lambda^{(N)} - \Lambda)hx\right\|$$

Similarly to Equation (7), the first term is of order $O(1/N)$. The second is bounded by $\|f - f^N\|\|Dh\| \leq L\|f - f^N\|$. □

3.4 Hölder Continuous Derivative

In this section, we study what happens when the derivative of the drift is not Lipschitz-continuous. When the drift f is Lipschitz-continuous with constant L, it is shown in [15] that there exists a constant C such that for all t:

$$\|\mathbb{E}[X^{(N)}(t)] - x(t)\| \leq \mathbb{E}[\|X^{(N)}(t) - x(t)\|] \leq C\frac{e^{Lt}}{\sqrt{N}}.$$

In the previous section, we have shown that when the derivative of the drift exists and is Lipschitz-continuous, the convergence is in $O(1/N)$. The next theorem, whose proof is postponed to Section 6.3, characterizes what happens in between, *i.e.* when the derivative of the drift exists and is α-Hölder continuous with $\alpha \in (0, 1)$.

Theorem 3.4. *Let $X^{(N)}$ be a population process on a Banach space $\mathcal{E}$ such that $\sup_x \sum_{\ell\in\mathcal{L}} \beta_\ell(x)\|l\|^2 < \infty$. Assume that the drift is Lipschitz-continuous with constant L and that its derivative exists and is α-Hölder continuous. Let $h : \mathcal{E} \to \mathbb{R}$ be a differentiable function such that h and its derivative Dh are α-Hölder continuous with constant L. Then, there exists $C < \infty$ such that for $t > 0$:*

$$\left|\mathbb{E}\left[h(X^{(N)}(t))\right] - h(\Phi_t x)\right| \leq CL\frac{e^{Ct}}{\sqrt{N}^{1+\alpha}}.$$

A key difference between this result and Theorem 3.1 lies in the methodology of proof. On the one hand, the proof of Theorem 3.4 involves Gronwall's lemma which leads to a constant e^{Ht} that grows exponentially with time which therefore cannot be easily adapted to the stationary regime. On the other hand, the proof of Theorem 3.1 involves the convergence of the generators. The constant in front of $1/N$ depends on the Lipschitz constant of $\int_0^t D\Phi_s ds$. This result can then be adapted to the the stationary measure by showing that the Lipschitz constant of $\int_0^\infty D\Phi_s ds$ is bounded.

In fact, the proofs of Theorem 3.1 and 3.2 can be adapted to the case of α-Hölder continuous derivative. The only difference is that, for a function g that has a α-Hölder continuous derivative, we have $(L^{(N)} - \Lambda)g = O(1/N^\alpha)$ instead of Equation (7). This means that the α-Hölder continuity of $\int_0^t D\Phi_s ds$ implies a convergence rate of $O(1/N^\alpha)$ (for $t < \infty$ and $t = +\infty$). As a consequence, if the ODE has a locally exponentially stable attractor x^*, if the derivative of the drift is α-Hölder continuous and if h equals 0 outside a bounded set, then in stationary regime we have

$$\mathbf{E}^{(N)}\left[h(X^{(N)}) - h(x^*)\right] \leq \frac{C}{N^\alpha} \tag{11}$$

More details are given in Sections 6.1 and 6.2.

To the best of our knowledge, Equation (11) is the first result that guarantees a rate of convergence of the steady-state distribution of a mean-field model in the case where the derivative is not Lipschitz-continuous ([26] considers models that evolve in a compact space and that have a twice-differentiable drift which implies that the derivative of the drift is Lipschitz-continuous).

The exponent α of Equation (11) is smaller than the exponent $(\alpha + 1)/2$ obtained in Theorem 3.4 for the transient regime. A natural question is whether the rate of convergence $O(1/\sqrt{N}^{1+\alpha})$ of Theorem 3.4 carries over to the stationary regime. We conjecture that it is the case. Yet, neither the method used to prove Theorem 3.2 nor the one used for Theorem 3.4 can be modified to show this. The former because it involves the convergence of the generator that needs an α-Hölder continuous derivative in order to get an $O(N^{-\alpha})$ convergence. The latter because the proof uses Gronwall's argument which leads to a constant e^{Ht} that explodes with time.

4 NECESSITY OF THE CONDITIONS

In this section, we examine the necessity of the regularity conditions of the drift. We provide an example that shows that when the drift is Lipschitz-continuous (and not differentiable), the rate of convergence of $|\mathbb{E}[X^{(N)}] - x|$ is not faster than $1/\sqrt{N}$. When the derivative of the drift is α-Hölder, we show that the convergence is not faster than $1/\sqrt{N}^{1+\alpha}$. This contrasts with the case of $\mathbb{E}[\|X^{(N)} - x\|]$ that is known to converge at rate $\Theta(1/\sqrt{N})$ in all cases.

4.1 Case-Study: a Birth-Death Process

We use a small example for which we are able to numerically solve the Kolmogorov equations associated to the N-dimensional Markov chain. This avoids estimation errors due to simulations.

Let $\alpha \in [0, 1]$. We consider a birth-death population process on $\{0, 1/N, \dots, 1\}$ whose drift is the function f_α:

$$f_\alpha(x) = \frac{1}{2} - x + \frac{1}{2}\left|\frac{1}{2} - x\right|^{1+\alpha}$$

The transition graph is represented in Figure 1. The transitions are as follows. For $x \in \{1/N, \dots, 1 - 1/N\}$, the chain jumps from state x to state $x + 1/N$ at rate $N + Nf(x)/2$ and from state x to state $x - 1/N$ at rate $N - Nf(x)/2$. The system jumps from state 0 to $1/N$ at rate $Nf(0)$ and from 1 to state $1 - 1/N$ at rate $-Nf(1)$.

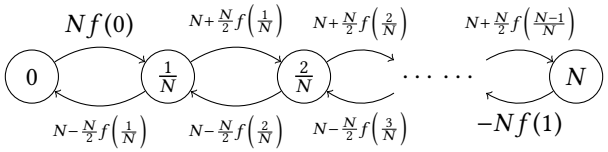


Fig. 1. Transition graph of the birth-death example.

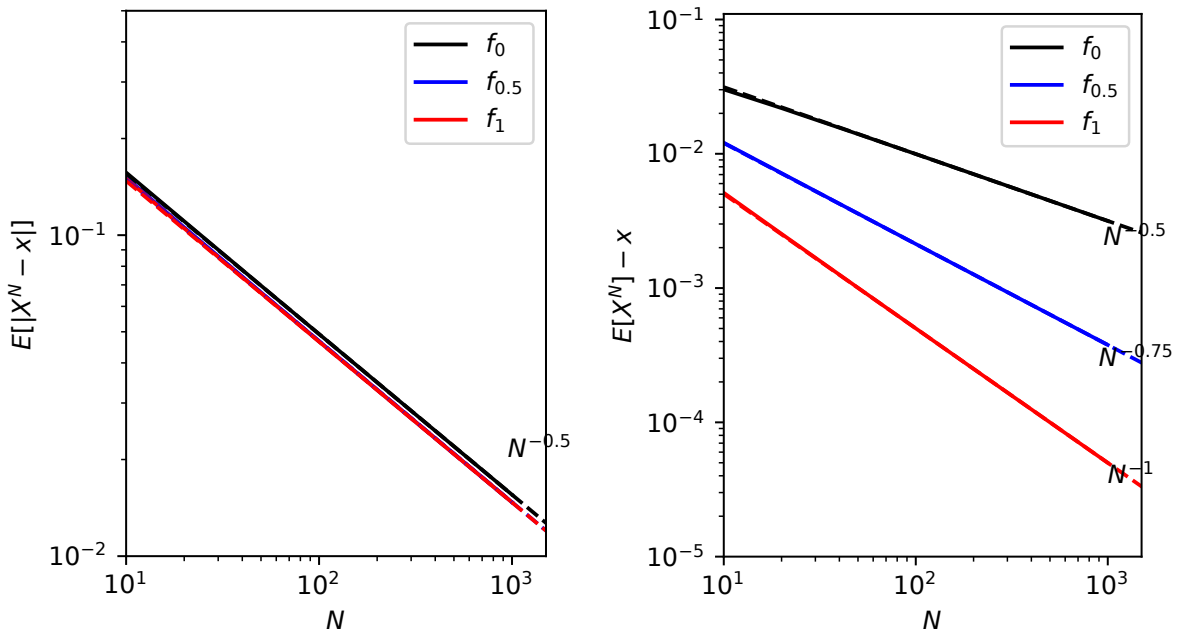
It should be clear that this process is a population process with drift f. For $\alpha = 0$, f is Lipschitz-continuous and not differentiable. For $\alpha = 1$, f is differentiable and its derivative is Lipschitz-continuous. For $\alpha \in (0, 1)$, the drift is differentiable and its derivative is α-Hölder continuous. In the remainder of this section, we study how α affects the convergence rate of $\mathbb{E}[X^{(N)}] - x$ and of $\mathbb{E}[\|X^{(N)} - x\|]$.

4.2 Differentiability of the Drift

When f is Lipschitz-continuous, $\mathbb{E}[\|X^{(N)} - x\|]$ converges at rate $O(1/\sqrt{N})$ as N goes to infinity [2, 15]. In general, the $O(1/\sqrt{N})$ is tight: It is the rate of convergence of the central limit theorem. We verify this in Figure 2(a) where we plot $\mathbb{E}[|X^{(N)}(1) - x(1)|]$ as a function of N with an initial state of $X^{(N)}(0) = 0.5$. We observe that it decreases as $O(1/\sqrt{N})$, (the axes are in log-scale). We also display a curve $c/\sqrt{N}$.

The situation of $\mathbb{E}[X^{(N)}] - x$ is different. Theorems 3.1 and 3.4 show that the convergence rate is at least $O(1/\sqrt{N}^{1+\alpha})$. In Figure 2(b), we plot as a function of N the difference $|\mathbb{E}[X^{(N)}] - x|$ for $\alpha \in \{0, 0.5, 1\}$. We also plot the function cN^{-e} that best fits the curve. In each case, the best fit value of the exponent is exactly $(1 + \alpha)/2$. This figure indicates that the exponent provided in Theorem 3.4 cannot be improved.

An important observation is that the Lipschitz-continuity of the drift is not sufficient to obtain an $O(1/N)$ convergence. This example shows that mean-field approximation is more accurate when the drift is not only Lipschitz-continuous but also has a derivative that is Lipschitz-continuous.



(a) $\mathbb{E}[\|X^{(N)}(1) - x(1)\|]$ (b) $|\mathbb{E}[X^{(N)}(1)] - x(1)|$

Fig. 2. Convergence rate in the case of the birth-death example. The Hölder exponent α has no impact on the rate of convergence of $\mathbb{E}[|X^{(N)} - x|]$ but does affect the rate of convergence of $|\mathbb{E}[X^{(N)}] - x|$.

4.3 Influence of the Hölder Exponent

In Figure 3, we study in more detail the impact of the exponent α on the rate of convergence in transient and stationary regime. In both cases, we observe that the exponent $(1+\alpha)/2$ given by Theorem 3.4 cannot be improved.

We used a numerical procedure to compute the values of $\mathbb{E}[|X^{(N)}(t) - x(t)|]$ and of $|\mathbb{E}[X^{(N)}(t)] - x(t)|$ for $\alpha \in [0, 1]$. The value for $t = 1$ is computed by integrating the Kolmogorov equation while the steady-state value is easily obtained because the process is a birth-death process. In all tested cases, we observed a convergence in cN^{-e}. In Figure 3, we report an estimation of the value of this exponent e as a function α. This figure confirms that the exponent of Theorem 3.4 is tight: For this model, the convergence occurs at rate $O(1/\sqrt{N}^{1+\alpha})$. Also, for this model, the exponent of the rate of convergence of the stationary distribution is the same as the one for the transient regime. Note again that the Hölder coefficient of the drift's derivative has no impact on the convergence rate of $\mathbb{E}[\|X^{(N)} - x\|]$, neither in the transient nor in the stationary regime.

5 APPLICATION : TWO-CHOICE MODEL

The two-choice paradigm is a well studied model since [20, 25]. In this section, we show that this model satisfies the assumptions of Theorems 3.1 and 3.2. We use these theorems to show that the marginal law and the average queue length in steady-state converge at rate $O(1/N)$ to their mean-field approximation. We also evaluate the constant of the $O(1/N)$ numerically and provide numerical evidence that it evolves as $1/(2-2\rho)$ as the load of the system ρ goes to 1. By using this, we develop an improved approximation that is $O(1/N^2)$-accurate.

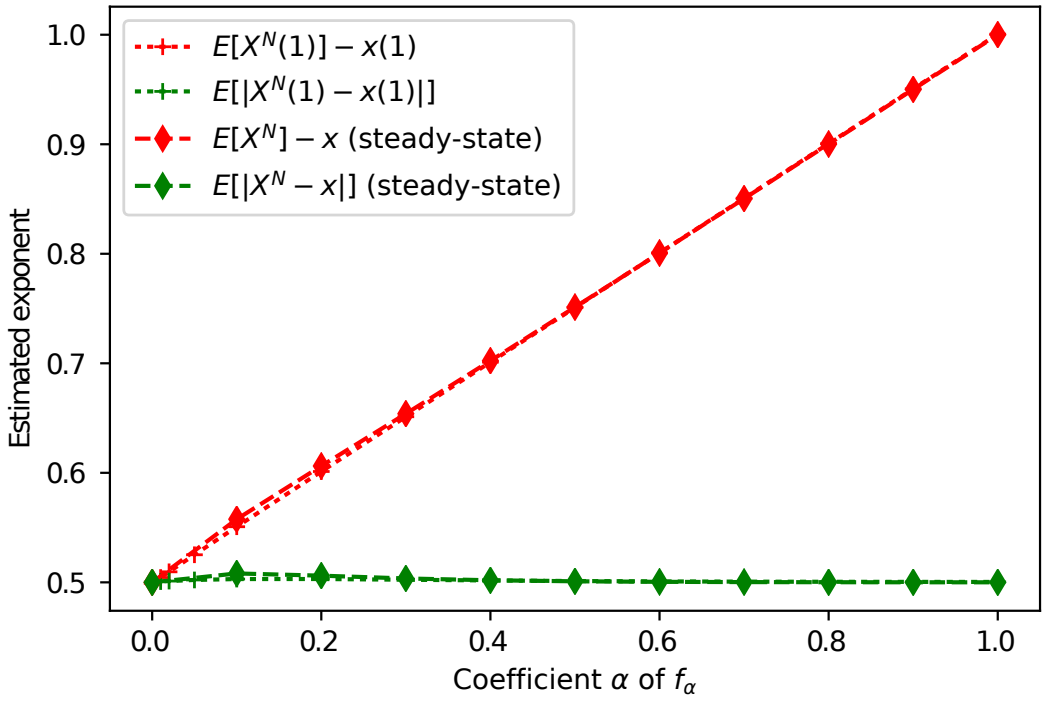


Fig. 3. Estimation of the exponent of the rate of convergence $O(N^{-\beta})$. For a drift that has a derivative that is α-Hölder continuous, this exponent is $(1+\alpha)/2$ for $\mathbb{E}[X^{(N)}] - x$ and is $1/2$ for $\mathbb{E}[\|X^{(N)} - x\|]$.

5.1 Two-choice Model

We consider the supermarket model of [20] with N identical servers. Each server maintains a separate queue. Jobs arrive in the system according to a Poisson process with intensity $N\rho$ and the processing time of a job is exponentially distributed with mean 1. We denote by $S_n^{(N)}(t)$ the queue size of the server n at time t. For each incoming job, two queues (say i and j) are selected uniformly at random and the job is allocated to the smallest queue among those two. Note that for all N, $(S_1^{(N)}(t) \dots S_N^{(N)}(t))$ is a CTMC and has a unique stationary distribution when $\rho < 1$.

Let $X_i^{(N)}(t)$ denotes the fraction of servers that have a queue size at least i at time t. Let $\mathbf{e}_i$ denote the infinite vector where the ith component equals 1, all others being equal to 0. $X^{(N)}$ is a CTMC with the following transitions (for all $i \geq 1$):

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

The first type of events correspond to a departure from a queue with i jobs. The second term corresponds to an arrival at a queue with $i-1$ jobs : An arrival is allocated to a queue with $i-1$ jobs if both queues have at least $i-1$ jobs and not both queues have i or more jobs.

The form of these transitions indicates that $X^{(N)}$ is a population process, as defined in Section 2. The corresponding mean-field approximation of this system is given by the following infinite system of ODEs:

$$\begin{aligned} \dot{x}_i &= \rho(x_{i-1}^2 - x_i^2) + x_{i+1} - x_i && \text{for } i \geq 1 \\ x_0 &= 1. \end{aligned}$$

It is shown in [20] that this ODE converges exponentially fast to its unique fixed point $\pi = (\pi_0, \pi_1, \dots)$, where

$$\pi_i = \rho^{2^i - 1}.$$

This guarantees that the stationary distribution of $X^{(N)}$ concentrates on π as N goes to infinity. In the following, we analyze the rate at which this concentration occurs.

5.2 Rate of Convergence

The only difficulty for applying the results of Section 3 is that the queue sizes are unbounded and that therefore, $X^{(N)}$ lives in an infinite-dimensional space. In the following, we equip the space of infinite sequences with a weighted ℓ_1-norm with a carefully chosen sequence w. We denote the norm of an infinite sequence x by $\|x\|_w$:

$$\|x\|_w = \sum_{i=1}^{\infty} w_i \, |x_i| \, .$$

Let $\mathcal{E}$ be the set of sequences x such that $\|x\|_w < \infty$. The process $X^{(N)}$ lives in a subset of $\mathcal{E}$ for each N: Indeed, the sequence $(X_i^{(N)})_i$ is non-negative, non-increasing, and as there is a finite number of jobs in each server, there is only a finite number of indices i for which $X_i^{(N)} > 0$. Such a sequence satisfies $\|X^{(N)}\|_w < \infty$ and is therefore in $\mathcal{E}$.

The next result, whose proof is given in Section 6.4, shows that there exists a sequence of weights w such that the model satisfies the assumptions of Theorem 3.2.

Theorem 5.1. *There exists an increasing sequence of weights $(w_i)_{i \geq 0}$ and a constant B such that the two-choice model satisfies the assumptions of Theorem 3.2 and such that in steady-state:*

$$\mathbf{E}^{(N)} \left[\left\| X^{(N)} \right\|_w \mathbf{1}_{\|X^{(N)}\|_w \geq B} \right] = o\left(\frac{1}{N}\right). \tag{12}$$

Theorem 3.2 implies that if h is a bounded function, then the quantity $\mathbf{E}^{(N)}[h(X^{(N)})]$ converges to $h(\pi)$ at rate $O(1/N)$. This is the case for the marginal law of one server by using the function $h(x) = x_i$ which is bounded by 1. This shows:

$$\mathbf{P}\left[\text{a server has } i \text{ or more jobs}\right] = \rho^{2^i - 1} + O\left(\frac{1}{N}\right).$$

This result sharpens the result of [18] in which a convergence rate of $O(\log N/N)$ was obtained.

Equation (12) implies that for any h that grows at most linearly, $\mathbf{E}^{(N)}[h(X^{(N)})]$ converges to $h(\pi)$ at rate $O(1/N)$.

Corollary 1. *Let $h : \mathcal{E} \to \mathbb{R}$ be a twice-differentiable function. Assume that there exists constants a, b such that for all $x \in \mathcal{E}$: $|h(x)| \leq a + b\|x\|_w$ for the sequence w defined in Theorem 5.1. Then the stationary measure of the N-server two-choice model satisfies*

$$\left| \mathbf{E}^{(N)} \left[h(X^{(N)}) \right] - h(\pi) \right| = O\left(\frac{1}{N}\right).$$

For a vector $X \in \mathcal{E}$, the quantity $N \sum_{i\geq 1} X_i$ is the total number of jobs in the N servers of the model. As the weight sequence w is increasing, the function $h(x) = \sum_i x_i$ satisfies $h(x) = \sum_i x_i \leq \|x\|_w / w_1$. As a result, the expected number of jobs satisfies

$$\mathbf{E}^{(N)}\left[\sum_{i\geq 1} X_i^{(N)}\right] = \sum_{i\geq 1} \rho^{2^i-1} + O\left(\frac{1}{N}\right). \tag{13}$$

Our result can also be used to bound the mean-square deviation by using $h(x) = \sum_{i=0}^{\infty}(x_i - \pi_i)^2$. As $x_i \leq 1$, this function is sub-linear: $h(x) \leq \sum_i x_i + \pi_i \leq (\|\pi\|_w + \|x\|_w)/w_1$. This shows that for the two-choice model:

$$\mathbf{E}^{(N)}\left[\sum_{i\geq 1}(X_i^{(N)} - \pi_i)^2\right] = O\left(\frac{1}{N}\right).$$

This last result improves the bound of [27], in which a convergence rate of $O((\log N)^3(\log\log N)^6/N)$ is obtained. The factor $(\log N)^3(\log\log N)^6$ of [27] comes from the fact that the author studies a truncated finite-dimensional model in order to study the original model. This factor is then a bound on the error between the finite and infinite models. One advantage of our approach is that we treat directly the infinite-dimensional system.

To work in the infinite-dimensional case, the main technical difficulty for applying our approach is to show that $\mathbf{P}[\|X^{(N)}\| \geq B] = O(1/N^2)$. In our proof of Theorem 5.1, we prove it by using a coupling to bound the process by a system of N independent $M/M/1$ queues. We then show in Lemma 6.5 that the sum of N independent variables have a third moment and therefore concentrate on a bounded set at rate $O(1/N^2)$.

5.3 Numerical Evaluation

In this section, we evaluate numerically the $O(1/N)$ of (13). We provide numerical experiments that show that when ρ is close to 1, the difference between the average queue length of the N-server model and the mean-field approximation is approximately $\rho^2/(2N(1-\rho))$.

5.3.1 Methodology. In order to evaluate the term $O(1/N)$, we use a stochastic simulator[3] of the two-choice model. Our simulator follows strictly the stochastic model that we described in Section 5.1.

To sample from the stationary distribution, we use a starting period of $10^4 N$ events starting from an empty system. We then draw one sample of a performance indicator by computing an average over the next 10^6 events. This gives us one sample. To guarantee confidence intervals of order $1/N$, the reported results are averages over $K(N) = \max(500, N^2/100)$ samples. The error bars that are shown on the figures correspond to the standard confidence intervals for the mean (*i.e.* $\pm 2\sigma/\sqrt{K(N)}$, where σ^2 is the empirical variance of our $K(N)$ samples). The time to simulate $K(N)$ samples grows as N^2: from 10 seconds for $N = 100$ to more than 1 hour for $N = 1000$.

5.3.2 Marginal law of one server. We first study the convergence of the marginal distribution of one object to π. For the system of size N, we denote by p_i^N the probability that a server has i or more jobs. As the servers are symmetric, p_i^N equals

$$p_i^N = \mathbf{P}\left[S_1^{(N)} \geq i\right] = \mathbf{E}^{(N)}\left[X_i^{(N)}\right].$$

By Theorem 5.1, the difference between p_i^N and π_i is of order $O(1/N)$. In Figure 4, we display the quantity $N(p_i^N - \pi_i)$ as a function of N. We report the values for $\rho = 0.9$ and $i \in \{2, 3, 4, 5\}$. The quantity $N(p_i^N - \pi_i)$ converges rapidly to a constant value.

[3]The code of our simulator and of all scripts needed to plot the graphs are available at https://github.com/ngast/meanFieldAccuracy.

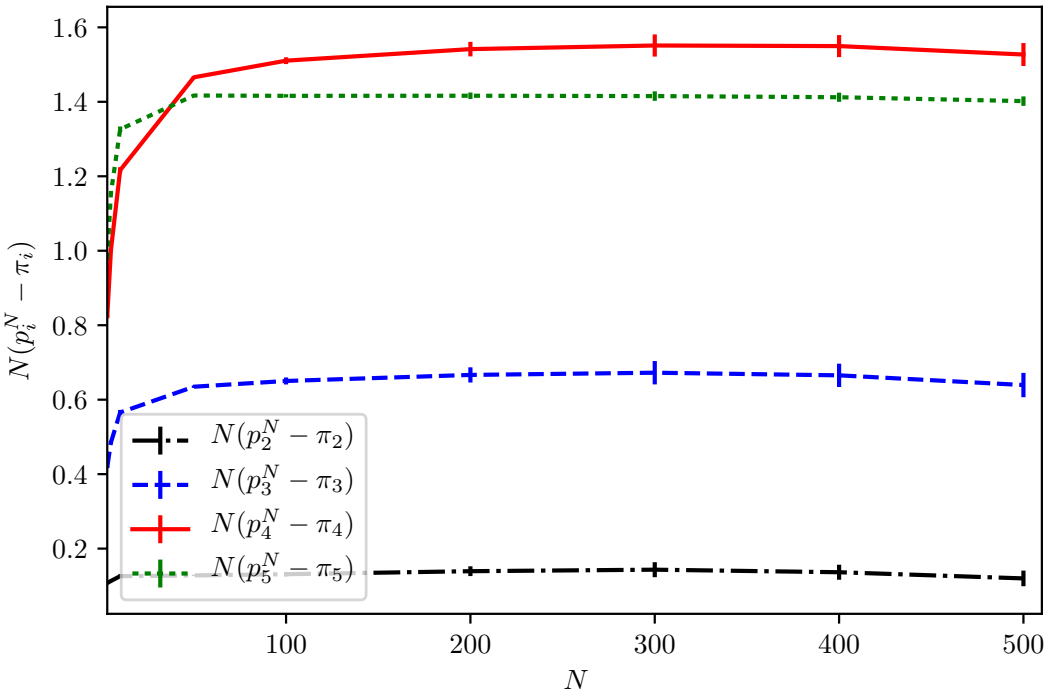


Fig. 4. Two-choice model: convergence of the marginal law of one server. We plot $N(p_i^N - \pi_i)$ as a function of N for various values of i and $\rho = 0.9$.

We also performed simulations for other values of $\rho \in [0.5, 0.99]$ for which we observe similar behavior. This suggests that for each $\rho \in [0, 1)$ and $i \geq 1$ there exists a constant $c_{i,\rho}$ such that

$$p_i^N = \pi_i + \frac{1}{N} c_{i,\rho} + o\left(\frac{1}{N}\right). \tag{14}$$

5.3.3 Average queue length. We now focus our attention on the average queue length (in steady-state). We denote by $m^N(\rho)$ the average queue length in the system of size N and by $m^\infty(\rho) = \sum_{i \geq 1} \pi_i$ its mean-field approximation.

By Corollary 1, $m^N(\rho) = m^\infty(\rho) + O(1/N)$. Equation (14) suggests that there exists a constant $d(\rho) = \sum_{i \geq 1} c_{i,\rho}$ such that

$$m^N(\rho) = m^\infty(\rho) + \frac{1}{N} d(\rho) + o\left(\frac{1}{N}\right). \tag{15}$$

We verify this numerically in Figure 5, where we display $N(m^N(\rho) - m^\infty(\rho))$ as a function of N for various values of ρ. We observe that in each case, this value stabilizes quickly. This stabilization is slower as ρ approaches 1.

Equation (15) shows that it is possible to obtain a more precise approximation by considering the mean-field approximation plus the $d(\rho)/N$ term. In Table 2, we study the accuracy of this improved approximation in the case $\rho = 0.9$. This table should be compared with Table 1 (shown in the introduction) that shows that the mean-field approximation is $1/N$-accurate. The improved approximation is indeed much closer to the value obtained by

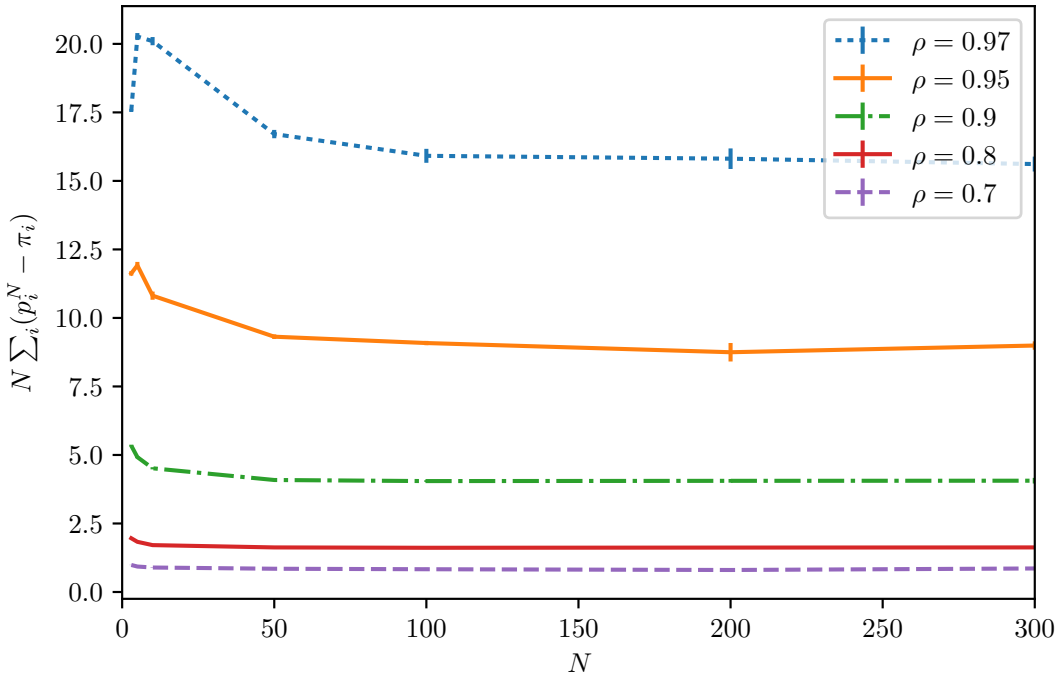


Fig. 5. Two-choice model: convergence of the average queue length m^N. We plot $N(m^N - m^\infty)$ as a function of N for various values of ρ.

N	10	20	30	50	$+\infty$
$m^N(\rho)$	2.804	2.566	2.491	2.434	–
$m^\infty(\rho)+d(\rho)/N$	2.759	2.556	2.488	2.434	2.353
Error	4.10^{-2}	1.10^{-2}	2.10^{-3}	3.10^{-4}	–

Table 2. Average queue length in the two-choice model (for $\rho = 0.9$). The values for a finite number of servers N are obtained by simulation. The error is the distance between m^N and the corrected mean-field approximation $m^\infty(\rho) + d(\rho)/N$. It is proportional to $1/N^2$.

simulation: We observe that the error of the improved approximation, $m^N(\rho) - m^\infty(\rho) - d(\rho)/N$, diminishes at rate $O(1/N^2)$. The $O(1/N^2)$ error term that we observe in Table 2 is the next term in the asymptotic expansion of $m^N(\rho)$, after the first term in $d(\rho)/N$.

Evolution of $d(\rho)$. As observed in Figure 5, the constant $d(\rho)$ increases rapidly as ρ approaches 1. To evaluate how $d(\rho)$ evolves with ρ, we plot in Figure 6(a) an estimate of the quantity $d(\rho)$ as a function of ρ. Our estimate of $d(\rho)$ is the experimental value for $N = 1000$, $1000(m^{1000}(\rho) - m^\infty(\rho))$. We also plot the function $\rho^2/(2-2\rho)$. We observe that this function fits very well the values for $\rho \in [0.5; 0.99]$. To confirm this, we also plot $d(\rho)(1-\rho)/\rho^2$ as a function of ρ in Figure 6(b). This figure suggests that it converges to $1/2$ as ρ goes to 1.

The average queue length of the mean-field limit has been shown in [20] to be of order $\log(1/(1-\rho))$ as ρ tends to 1: more precisely, $\lim_{\rho\to 1} m^{\infty}(\rho)/\log(1/(1-\rho)) = 1/\log 2$. This is an order of magnitude smaller than when jobs are allocated at random in which case the average queue length is $1/(1-\rho)$. Our numerical result suggests that for a finite system size N, a term in $O(1/(N(1-\rho)))$ remains.

We conjecture that the average queue length of a N-server two-choice system satisfies

$$m^N(\rho) = \Theta_{\rho\to 1}\left(\log\frac{1}{1-\rho}\right) + \frac{1}{N}\Theta_{\rho\to 1}\left(\frac{1}{1-\rho}\right) + O\left(\frac{1}{N^2}\right).$$

with a hidden constant of 1/2 in the term $\Theta_{\rho\to 1}(1/(1-\rho))$. This term is equal to $d(\rho)$ and seems to satisfy

$$\lim_{\rho\to 1}(1-\rho)d(\rho) = \frac{1}{2}.$$

6 PROOFS

Notations

For a function $f : \mathcal{E} \to \mathcal{F}$ and $g : \mathcal{F} \to \mathcal{G}$, we denote by $Df(x)$ the derivative of the function f evaluated at x. $Df(x)$ is a linear map that associates to each u the quantity $Df(x)\cdot u$. We denote by $D(f(g(x)))$ the derivative of the function $f\circ g$ with respect to x evaluated at x and by $Df(g(x))$ the derivative of f evaluated at $g(x)$. It follows that $D(f(g(x))) = Df(g(x))\cdot Dg(x)$.

We define the following operators $\Psi_t^{(N)}$ and Φ_t that operate on $h : \mathcal{E} \to \mathbb{R}$:

$$\Psi_t^{(N)}h(x) = \mathbb{E}\left[h(X^{(N)}(t)) \mid X^{(N)}(0) = x\right] \tag{16}$$

$$\Phi_t h(x) = h(\Phi_t x) \tag{17}$$

For a function $h \in C^1(\mathcal{E},\mathbb{R})$, we have $\frac{d}{dt}\Psi_t^{(N)}h = \Psi_t^{(N)}L^{(N)}h = L^{(N)}\Psi_t^{(N)}h$ and $\frac{d}{dt}\Phi_t h = \Phi_t\Lambda h = \Lambda\Phi_t h$.

6.1 Proof of Theorem 3.1 (Transient Regime)

The proofs of the following lemmas follow similar ideas as the proof of [14, Theorem 1]. Yet, we provide a complete proof here as our result generalizes [14, Theorem 1] for two reasons. First, the set of transitions that we consider is more general (a transition in the model of [14] can only affect one object). Second, our results apply for Banach space and not just for finite dimensional sets.

Lemma 6.1. *For any $h \in C^1(\mathcal{E},\mathbb{R})$, we have*

$$\mathbb{E}\left[\left|h(X^{(N)}(t)) - h(\Phi_t x)\right|\right] \le \int_0^t \sup_{x\in\mathcal{E}} |(\Lambda - L^{(N)})h(\Phi_s x)|ds. \tag{18}$$

Proof. By definition of the operators $\Psi^{(N)}$ and Φ introduced in (16) and (17), the left-hand side of Equation (18) can be written as

$$\mathbb{E}\left[h(X^{(N)}(t)) - h(\Phi_t x)\right] = \Psi_t^{(N)}hx - \Phi_t hx = (\Psi_t^{(N)} - \Phi_t)hx.$$

We are therefore interested in the convergence of $\Psi_t^{(N)} - \Phi_t$.

As in [14, Theorem 1], we use the following trick : The function $s \mapsto \Psi_{t-s}^{(N)}\Phi_s$ is equal to Φ_t for $s = t$ and $\Psi_t^{(N)}$ for $s = 0$. Hence:

$$\begin{aligned}\Phi_t - \Psi_t^{(N)} &= \int_0^t \frac{d}{ds}\Psi_{t-s}^{(N)}\Phi_s ds \\ &= \int_0^t \Psi_{t-s}^{(N)}(\Lambda - L^{(N)})\Phi_s ds.\end{aligned} \tag{19}$$

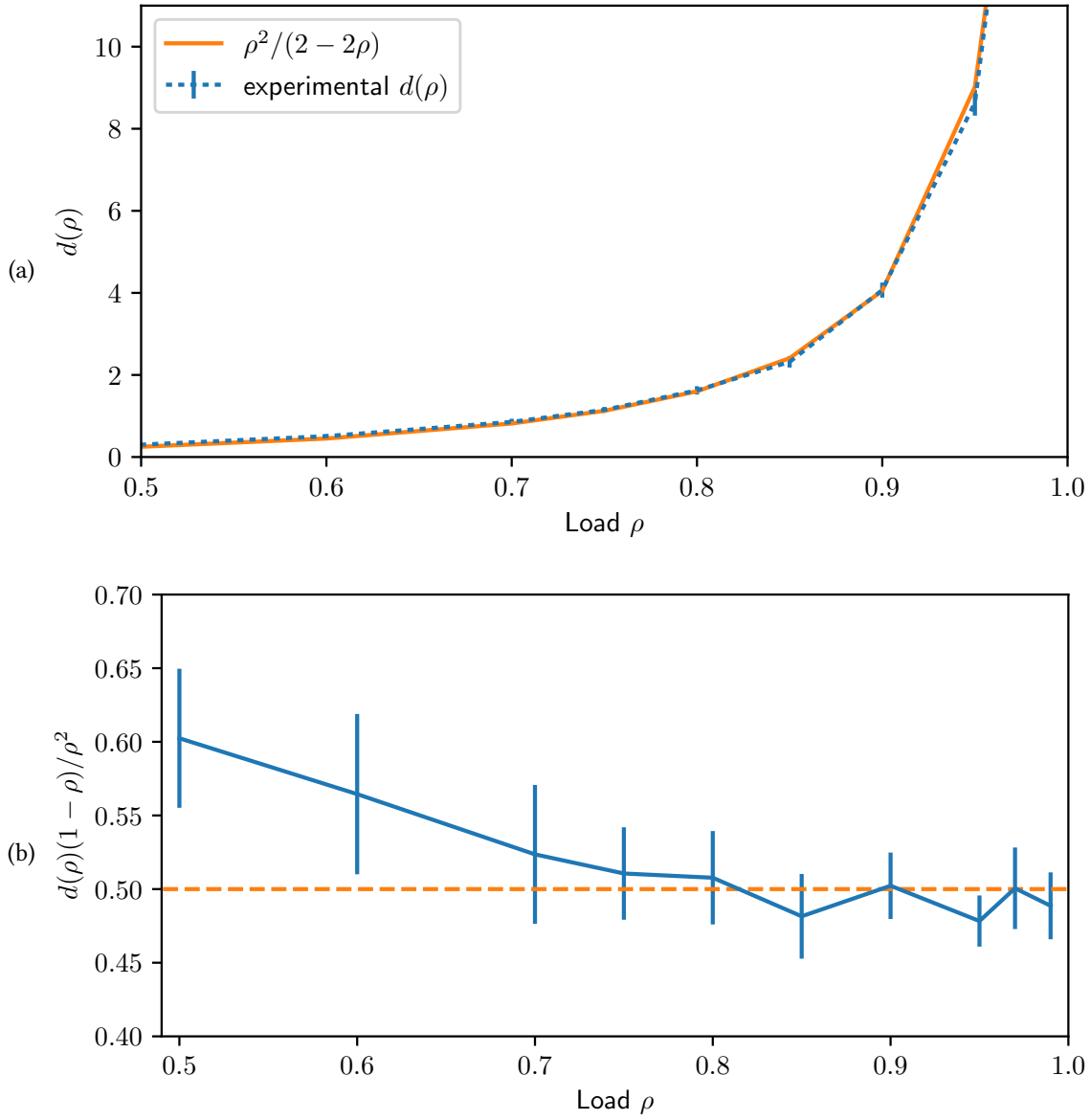


Fig. 6. Evaluation of $d(\rho)$ as a function of ρ. In (a), we plot the two curves $d(\rho)$ and $\rho^2/(2-2\rho)$ that are almost indistinguishable. In (b), we plot $d(\rho)(1-\rho)/\rho^2$ which seems to go to $1/2$ as ρ goes to 1.

For any function h', one has

$$\left|\Psi_{t-s}^{(N)} h' x\right| = \left|\mathbf{E}^{(N)}\left[h'(X^{(N)}(t-s)) \mid X^{(N)}(0) = x\right]\right| \leq \sup_{x\in\mathcal{E}} |h'(x)|\,.$$

This shows that for any h and any x:

$$\left|\Psi_t^{(N)} hx - \Phi_t hx\right| \leq \int_0^t \sup_{x\in\mathcal{E}} \left|(\Lambda - L^{(N)})\Phi_s hx\right| ds,$$

which is Equation (18). □

Remark: *Although Theorem 3.1 is stated only in the Lipschitz-continuous case, the proof in this section holds for α-Hölder continuous. This does not require more work and will be useful for the proof of Equation (11).*

Lemma 6.2. *Assume that Df is bounded and is $\alpha-$Hölder continuous with constant L. Then, there exists a constant β such that, for any t, Φ_t is α-Hölder continuous with constant $Le^{\beta t}$.*

Proof. Let $x, e \in \mathcal{E}$. By [4, Theorem 3.7.1 of Part II], $D\Phi_t(x)$ and $D\Phi_t(x+e)$ exist and are differentiable with respect to time. Moreover, $\frac{d}{dt}D\Phi_t = D(f(\Phi_t)) = Df(\Phi_t)\cdot D\Phi_t(x)$. As $\|D\Phi_0\| = 1$, this shows $\|D\Phi_t\| \leq \exp(t\|Df\|)$.

Let $F_t = D\Phi_t(x+e) - D\Phi_t(x)$. We have

$$\begin{aligned}
\frac{d}{dt}F_t &= \frac{d}{dt}\left(D\Phi_t(x+e) - D\Phi_t(x)\right)\\
&= D(f(\Phi_t(x+e))) - D(f(\Phi_t))\\
&= Df(\Phi_t(x+e))\cdot D\Phi_t(x+e) - Df(\Phi_t x)\cdot D\Phi_t(x)\\
&= Df(\Phi_t(x+e))\cdot F_t\\
&\quad - \left(Df(\Phi_t(x+e)) - Df(\Phi_t x)\right)\cdot D\Phi_t(x) \qquad (20)
\end{aligned}$$

By using the Hölder-continuity of Df and the fact that $\|D\Phi_t\| \leq e^{t\|Df\|}$, the second term of Equation (20) is bounded by

$$\begin{aligned}
\|Df(\Phi_t(x+e)) - Df(\Phi_t x)\| &\leq L\|D\Phi_t(x+e) - Df(\Phi_t x)\|^\alpha\\
&\leq L(e^{t\|Df\|}\|e\|)^\alpha
\end{aligned}$$

The quantity $\|F_t\|$ can then be bounded by using Gronwall's Lemma:

$$\begin{aligned}
\|F_t\| &\leq \int_0^t (\|Df\|\|F_s\| + L(e^{t\|Df\|}\|e\|)^\alpha)ds\\
&\leq Le^{t(1+\alpha)\|Df\|}\|e\|^\alpha.
\end{aligned}$$

The result follows by setting $\beta = (1+\alpha)\|Df\|$. □

6.2 Proof of Theorem 3.2 and of Equation (11) (Stationary Regime)

Remark: *As for Lemma 6.2, we directly prove the result in the case of α-Hölder continuous derivative. Theorem 3.2 can be obtained by considering the special case $\alpha = 1$.*

Lemma 6.3. *Let Φ_t be the flow associated to a drift $f : \mathcal{E} \to \mathcal{E}$ that is differentiable with a bounded derivative. Let $\mathcal{B} \subset \mathcal{E}$ be a bounded set. Assume that f is α-Hölder continuous and that the flow has an exponentially stable attractor x^*.*

(i) *There exists $b, \delta > 0$ such that for any t:*

$$\sup_{x\in\mathcal{B}} \|D\Phi_t\| \leq be^{-\delta t}.$$

(ii) *There exists a constant c such that, the derivative of $x \mapsto \Phi_t x$ exists and is α-Hölder continuous on $\mathcal{B}$ with constant $ce^{-\delta t}$.*

(iii) *The function Gh, defined by*

$$Gh : x \mapsto \int_0^\infty (h(\Phi_t x) - h(x^*))\, dt,$$

is differentiable on $\mathcal{B}$ and its derivative is α-Hölder continuous.

Proof. We will first prove *(i)*, then prove that *(i)* implies *(ii)* and then prove *(iii)*. The most technical part of our proof is to show that *(i)* implies *(ii)*. Its proof is inspired by the proof of [8, Lemma C.1] that shows that the exponential decay of the first derivative of the flow implies the exponential decay of the second derivative of a flow.

Proof of *(i)* – To ease the notation, we assume without loss of generality that $x^* = 0$. By assumption, 0 is exponentially stable, which means that there exists $\beta > 0$ such that $\|\Phi_t(x)\| \le e^{-\beta t}\|x\|$.

Let $A = Df(0)$. As Df is continuous at $x^* = 0$, $f(x) = Ax + g(x)$ and $Df(x) = A + h(x)$ where the functions g and h satisfy $\lim_{x\to 0} \|g(x)\|/\|x\| = 0$ and $\lim_{x\to 0} \|h(x)\| = 0$. Let $\delta = \beta/2$. By assumption, there exists a neighborhood $\mathcal{N}$ of 0 such that $\|g(x)\| \le \delta\|x\|$ and $\|h(x)\| \le \delta$ for $x \in \mathcal{N}$. As the flow is stable, there exists a neighborhood $\mathcal{N}'$ of 0 such that if $x \in \mathcal{N}'$, then $\Phi_t x \in \mathcal{N}$ for all $t \ge 0$.

The flow satisfies

$$\Phi_t x = e^{At}x + \int_0^t e^{A(t-s)} g(\Phi_s x) ds$$

As the flow is exponentially stable, it is also linearly stable: $\|e^{At}\| \le e^{-\beta t}$. Hence, for all $x, x+y \in \mathcal{N}'$:

$$\begin{aligned}
&\|\Phi_t(x+y) - \Phi_t x\| \\
&= \left\| e^{At}y + \int_0^t e^{A(t-s)}(g(\Phi_s(x+y)) - g(\Phi_s x))ds \right\| \\
&\le e^{-\beta t}\|y\| \\
&\qquad + \int_0^t e^{-\beta(t-s)} \underbrace{\sup_{x\in N_\varepsilon} \|h(x)\|}_{\le\delta} \|\Phi_s(x+y) - \Phi_s(x)\| ds.
\end{aligned}$$

Therefore, $z(t) = \|\Phi_t(x+y) - \Phi_t x\| e^{\beta t}$ satisfies

$$z(t) \le \|y\| + \delta \int_0^t z(s)ds \le \sup_{y\in\mathcal{N}'} \|y\| e^{\delta t}$$

This shows that Φ_t is Lipschitz-continuous on $\mathcal{N}'$ with a constant less than $ae^{-\delta t}$ with $a = \sup_{y\in\mathcal{N}'} \|y\|$. By Lemma 6.2, $D\Phi_t$ is continuous, which implies that $\|D\Phi_t\| \le ae^{-\delta t}$.

The set $\mathcal{B}$ is a bounded subset of $\mathcal{E}$ and 0 is exponentially stable. Hence, there exists T such that for all $x \in \mathcal{B}$, we have $\Phi_t(x) \in \mathcal{N}'$. Let $x \in \mathcal{B}$. For $t \ge T$, we have

$$\begin{aligned}
\|D\Phi_t x\| = \|D(\Phi_{t-T} \circ \Phi_T)\| &= \|D\Phi_{t-T}(\Phi_T) \cdot D\Phi_T\| \\
&\le a^{-\delta(t-T)} e^{\|Df\|T} \le be^{-\delta t},
\end{aligned}$$

with $b = a^{\delta T + \|Df\|T}$.

Proof of (i)⇒*(ii)* – As in the proof of Lemma 6.2, let $F_t = D\Phi_t(x+e) - D\Phi_t(x)$. Following, Equation (20), F_t satisfies the differential equation

$$\frac{d}{dt}F_t = Df(\Phi_t(x+e)) \cdot F_t \\ - (Df(\Phi_t(x+e)) - Df(\Phi_t x)) \cdot D\Phi_t(x) \tag{21}$$

with initial condition $F_0 = D(\Phi_0(x+e)) - D\Phi_0(x) = 0$.

The homogeneous part of the differential equation is $\dot{A} = Df(\Phi_t(x+e))A$. It is a linear differential equation on $\mathcal{L}(\mathcal{E},\mathcal{E})$, the set of linear map from $\mathcal{E}$ to $\mathcal{E}$. Let $A_{t,s} : \mathcal{L}(\mathcal{E},\mathcal{E}) \to \mathcal{L}(\mathcal{E},\mathcal{E})$ be the solution operator[4] of this homogeneous differential equation.

Let $F : \mathbb{R}^+ \to \mathcal{L}(\mathcal{E},\mathcal{E})$ be the function defined by

$$F_t = \int_0^t A_{t,s}\left(Df(\Phi_s(x+e)) - Df(\Phi_s x)\right) \cdot D\Phi_s x ds.$$

Let us show that F is indeed the function F defined at the beginning of the proof by showing that it is a solution of the differential equation (21). By definition of F, we have $F_0 = 0$. Moreover, for $t > 0$, we have

$$\begin{aligned}\frac{d}{dt}F_t &= \int_0^t \frac{d}{dt}A_{t,s}\left(Df(\Phi_s(x+e)) - Df(\Phi_s x)\right) \cdot D\Phi_s x ds \\ &\quad + A_{t,t}\left(Df(\Phi_t(x+e)) - Df(\Phi_t x)\right) \cdot D\Phi_t x \\ &= Df(\Phi_t(x+e)) \cdot F_t \\ &\quad + (Df(\Phi_t(x+e)) - Df(\Phi_t x)) \cdot D\Phi_t x.\end{aligned}$$

The solution operator A satisfies $A_{t,s}\mathbf{1} = D\Phi_{t-s}x$. Hence, by using *(i)* that $\|D\Phi_t\| \le ae^{-\delta t}$, we have $\|A_{t,s}\| \le ae^{-\delta(t-s)}$. This shows that

$$\begin{aligned}\|F_t\| &\le \int_0^t \|A_{t,s}\|\|Df(\Phi_s(x+e)) - Df(\Phi_s(x))\|\|D\Phi_s x\| ds \\ &\le \int_0^t ae^{-\delta(t-s)}L\|\Phi_s(x+e) - \Phi_s x\|^\alpha ae^{-\delta s} ds,\end{aligned} \tag{22}$$

where L is the Hölder constant of Df.

By using again *(i)* ($\|D\Phi_t\| \le ae^{-\delta t}$), we have

$$\|\Phi_s(x+e) - \Phi_s x\| \le \|\Phi_s\|\|e\| \le ae^{-\delta s}\|e\|$$

Plugging this bound into Equation (22) shows

$$\begin{aligned}\|F_t\| &\le \int_0^t ae^{-\delta(t-s)}L(ae^{-\delta s})^\alpha ae^{-\delta s} ds \|e\|^\alpha \\ &= La^{2+\alpha}e^{-\delta t}\int_0^t e^{-\delta\alpha s} ds \|e\|^\alpha \\ &= La^{2+\alpha}e^{-\delta t}\frac{1}{\delta\alpha}(1 - e^{-\delta\alpha t})\|e\|^\alpha\end{aligned}$$

This shows that $\|F_t\| \le ce^{-\delta t}\|e\|^\alpha$ with $c := La^{2+\alpha}/(\delta\alpha)$.

[4]This means that for $y \in \mathcal{L}(\mathcal{E},\mathcal{E})$, the function $s \mapsto A_{t,s}y$ is the solution of the homogeneous differential equation starting in y at time s.

Proof of *(iii)* – Lemma 6.2 shows that for all t, the derivative of Φ_t exists and is α-Hölder continuous. By *(i)*, the derivative of Gh exists and is equal to

$$D(Gh) = \int_0^\infty D(\Phi_t h)dt.$$

By *(ii)*, $D(\Phi_t h)$ is α-Hölder continuous with constant $Lce^{-\delta t}$. This shows that the function $D(Gh)$ is α-Hölder continuous with constant $\int_0^\infty Lce^{-\delta t}dt = Lc/\delta$. □

6.2.1 *Dealing with non-compact state-space.*

Lemma 6.4. *Assume that the system satisfies the assumption of Theorem 3.2 and the function h equals $h(x^*)$ outside a bounded set. Then there exists a constant B' such that*

$$\mathbf{E}^{(N)}\left[(\Lambda - L^{(N)})Gh(X^{(N)})\mathbf{1}_{X^{(N)}\notin\mathcal{B}}\right] < B'/N.$$

Proof. Let W be a bounded set outside containing the set V of Assumption (A2) and outside which $h(x)$ is equal to $h(x^*)$. For $x \in \mathcal{E}$, let $S(x)$ be the first time at which the solution of the ODE that starts in x enters W – the function $S(x)$ exists because $V \subset W$ and V attracts all trajectories. We have

$$\begin{aligned} Ghx &= \int_0^\infty (h(\Phi_t x) - h(x^*))dt \\ &= \int_0^{S(x)} \underbrace{(h(\Phi_t x) - h(x^*))}_{=0}\, dt + \int_{S(x)}^\infty (h(\Phi_t x) - h(x^*))dt \\ &= Gh(\Phi_{S(x)}x). \end{aligned}$$

As W is bounded and Gh is continuous, there exists $v < \infty$ such that $\sup_{x\in W}\|Ghx\| \le v < \infty$. Moreover, by definition of $S(x)$, we have $\Phi_{S(x)}x \in W$ which implies that for all x, $\|Ghx\| \le v$. This shows that

$$\begin{aligned} (L^{(N)}Gh)(x) &= \left\| \sum_{\ell\in\mathcal{L}} N\beta_\ell(x)\left((Gh)(x+\frac{\ell}{N}) - (Gh)(x)\right)\right\| \\ &\le N \sup_{x'\in\mathcal{E}} \sum_{\ell\in\mathcal{L}} \beta_\ell(x')2v \end{aligned}$$

Let $b := 2v\sup_{x'\in\mathcal{E}}\sum_{\ell\in\mathcal{L}}\beta_\ell(x') + 2\sup_x\|h(x)\| < \infty$. By using that $\Lambda(Gh)(x) = h(x) - h(x^*)$, we have

$$\left\|(\Lambda - L^{(N)})(Gh)(x)\right\| \le bN$$

This shows that

$$\mathbf{E}^{(N)}\left[(\Lambda - L^{(N)})(Gh)(X^{(N)})\mathbf{1}_{X^{(N)}\notin\mathcal{B}}\right] \le \mathbf{E}^{(N)}[Nb],$$

which is of order $O(1/N)$ by Assumption (A3). □

6.3 Proof of Theorem 3.4 (Hölder-continuity)

To simplify the notation, we assume that $X^{(N)}(0) = x$.

By [16], $X^{(N)}$ is equal to

$$X^{(N)}(t) = X^{(N)}(0) + \int_0^t f(X^{(N)}(s))ds + M_t,$$

where M_t is a martingale (in particular $\mathbb{E}[M_t] = 0$) that satisfies $\mathbb{E}[\|M_t\|^2] \leq C_t^2/N$ for some constant $C_t > 0$. Let $Y^{(N)}(t) = X^{(N)}(t) - \Phi_t x$. As $\Phi_t x = x + \int_0^t f(\Phi_s x)ds$ and $Y^{(N)}(0) = 0$, the quantity $Y^{(N)}$ satisfies

$$Y^{(N)}(t) = \int_0^t f(\Phi_s x + Y^{(N)}(s)) - f(\Phi_s x)ds + M_t. \tag{23}$$

Let H be the max between the Lipschitz-constant of f and the Hölder-constant of Df. By Gronwall's Lemma [4, Lemma 1.5.1, Part II]

$$\mathbb{E}\left[\left\|Y^{(N)}(t)\right\|^2\right] \leq \frac{2C_t^2 e^{2H^2 t}}{N}.$$

As Df is α-Hölder continuous, for any $x, y \in \mathcal{E}$ we have $\|f(x+y) - f(x) - Df(x)\cdot y\| \leq H\|y\|^{1+\alpha}$. As a result, for a random variable Y, we have

$$\mathbb{E}\left[\|f(x+Y) - f(x)\|\right] \leq \|Df\|\|\mathbb{E}[Y]\| + H\mathbb{E}\left[\|Y\|^{1+\alpha}\right]$$

Applying this to Equation (23) shows that

$$\left\|\mathbb{E}\left[Y^{(N)}(t)\right]\right\| \leq \int_0^t \|DF\|\left\|\mathbb{E}\left[Y^{(N)}(s)\right]\right\|ds + \frac{HT2^{(1+\alpha)/2}e^{H^2 t(1+\alpha)}}{N^{(1+\alpha)/2}},$$

By Gronwall's lemma [4, Lemma 1.5.1, Part II], this in turns implies the existence of a constant c such that $\|\mathbb{E}[Y^{(N)}(t)]\| \leq ce^{ct}N^{-(1+\alpha)/2}$.

6.4 Proof of Theorem 5.1 (Two-choice)

To prove that the two-choice model satisfies the assumptions of Theorem 3.2, the only difficulty is to show that the fixed point is asymptotically stable (assumptions A1 and A2) and to show that the stationary measure concentrates on a bounded sub-set of $\mathcal{E}$ (assumptions A3). The existence of a sequence of w such that the two-choice model is asymptotically stable for the weighted ℓ_1-norm $\|x\| = \sum_i w_i x_i$ has been shown in [20, Theorem 3.6]. To prove our concentration result, we also need that this sequence grows slower than $\rho^{-i/3}$. We construct such a sequence below.

According to [20, Proof of Theorem 3.6], the quantity $V(x) = \|x - \pi\|_w$ satisfies $\frac{d}{dt}V(\Phi_t x) \leq -\delta V(x)$ if the sequence w_i is positive and satisfies

$$w_{i+1} \leq w_i + \frac{w_i(1-\delta) - w_{i-1}}{\rho(2\pi_i + 1)}, \tag{24}$$

where $\pi_i = \rho^{2^i - 1}$ is the fixed point of the ODE.

Let $a = \rho^{-1/4}$ and let us show that there exists a sequence (w_i) and a constant $\delta > 0$ that satisfy (24) and such that $w_i = O(a^i)$.

Let $b \in (a, 1/\rho)$. As $b < 1/\rho$, there exists i^* such that $1/(\rho(2\pi_i + 1)) \geq b$ for $i \geq i^*$. We define a sequence w_i by $w_0 = 0$, $w_1 = 1$ and

$$w_{i+1} = w_i + \frac{w_i(1-\delta) - w_{i-1}}{3\rho} \qquad \text{for } i < i^*$$
$$w_{i+1} = aw_i \qquad \text{for } i \geq i^*.$$

By definition of w, $w_i = O(a^i)$.

As i^* is finite, there exists δ^* such that for $\delta \le \delta^*$, the sequence w_i is increasing and satisfies (24) for $i < i^*$. Let $\delta = \min(\delta^*, (a-1)(b-a)/(ab))$. For $i > i^*$, we have

$$\begin{aligned} w_{i+1} - w_i - \frac{w_i(1-\delta) - w_{i-1}}{\rho(2\pi_i + 1)} \\ &\le w_{i+1} - w_i - b(w_i(1-\delta) - w_{i-1}) \\ &= a(w_i - w_{i-1}) - b(w_i - w_{i-1}) + \delta b w_i \\ &= (a-b)(w_i - w_{i-1}) + \delta b w_i \\ &= (a-b)\left(1 - \frac{1}{a}\right) w_i + \delta b w_i \\ &= \left(\frac{(a-b)(a-1)}{a} + \delta b\right) w_i, \end{aligned}$$

this shows that w_i satisfies Equation (24) for $\delta = \min(\delta^*, (b-a)(a-1)/(ab)) > 0$.

According to [23, Theorem 4] (see also [13, Theorem 4.1]), there exists a coupling between $X^{(N)}$ and a variable $\tilde{X}^{(N)}$ picked according to the stationary distribution of N independent M/M/1 queues such that for all j, $\sum_{i=j}^{\infty} X_i^{(N)} \le \sum_{i=j}^{\infty} \tilde{X}_i^{(N)}$. Denoting $w_0 = 0$, for any infinite sequence x we have $\sum_{i=1}^{\infty} w_i x_i = \sum_{i=1}^{\infty}(w_i - w_{i-1}) \sum_{j=i}^{\infty} x_j$. As the sequence w is increasing, $w_i - w_{i-1} \ge 0$. This implies that (for this coupling): $\|X^{(N)}\|_w = \sum_{i=1}^{\infty} w_j X_j^{(N)} = \sum_{i=1}^{\infty}(w_i - w_{i-1}) \sum_{j=i}^{\infty} X_j^{(N)} \le \sum_{i=1}^{\infty}(w_i - w_{i-1}) \sum_{j=i}^{\infty} \tilde{X}_j^{(N)} = \|\tilde{X}^{(N)}\|_w$.

In stationary regime, the quantity $\tilde{X}^{(N)}$ can be written as $\tilde{X}_i^{(N)} = (1/N) \sum_{n=1}^{N} 1_{\tilde{Z}_n \ge i}$ where the variables $\tilde{Z}_n$ are *i.i.d.* and are geometrically distributed with parameter ρ. The sequence w defined above satisfies $w_i \le c a^i$. This implies that $\sum_{j=1}^{i} w_i \le c' a^i$, where $c' = ca/(a-1)$. Therefore, in the stationary regime, we have

$$\left\|X^{(N)}\right\|_w \le \left\|\tilde{X}^{(N)}\right\|_w = \frac{1}{N} \sum_{n=1}^{N} \sum_{i \ge 0} w_i \mathbf{1}_{\tilde{Z}_n \ge i} \le \frac{c'}{N} \sum_{n=1}^{N} a^{\tilde{Z}_n},$$

where $c' < \infty$ is the constant defined above.

To conclude, we use the following concentration lemma that we apply with $\xi_n = a^{\tilde{Z}_n}$. Note that quantity $\mathbb{E}[\|\xi_n\|^3] = \mathbb{E}[a^{3\tilde{Z}_n}]$ is finite because $a < \rho^{-1/3}$ and $\tilde{Z}_n$ is geometrically distributed with parameter ρ.

Lemma 6.5. *Let (ξ_n) be a sequence of* i.i.d. *random variables such that $\mathbb{E}[\|\xi_n\|^3] < +\infty$. Let $S_N = N^{-1} \sum_{n=1}^{N} \xi_n$. Then there exist two finite constants B and B' such that $\mathbb{E}[\|S_N\| \mathbf{1}_{\|S_N\| \ge B'}] \le B/N^2$.*

Proof. Let $\bar{\xi}_n = \xi_n - \mathbb{E}[\xi_n]$ and $\bar{S}_N = S_N - \mathbb{E}[S_N] = S_N - \mathbb{E}[\xi_1]$. By assumption, there exists a constant $B < \infty$ such that $\mathbb{E}[\|\bar{\xi}_n\|^3] \le B$. As the variables $\bar{\xi}_n$ are *i.i.d*, this implies that $\mathbb{E}[\|\bar{S}_N\|^3] \le B/N^2$. This shows that

$$\begin{aligned} \mathbb{E}\left[\left\|\bar{S}_N\right\| \mathbf{1}_{\|S_N\| \ge 1}\right] &\le \mathbb{E}\left[\left\|\bar{S}_N\right\|^3 \mathbf{1}_{\|S_N\| \ge 1}\right] \\ &\le \mathbb{E}\left[\left\|\bar{S}_N\right\|^3\right] \le \frac{B}{N^2}. \end{aligned}$$

By taking $B' = \mathbb{E}[\xi_1] + 1$, we have

$$\begin{aligned} \mathbb{E}\left[\|S_N\| \mathbf{1}_{\|S_N\| \ge B'+1}\right] &\le B' \mathbb{P}\left(\left\|\bar{S}_N\right\| \ge 1\right) + \mathbb{E}\left[\left\|\bar{S}_N\right\| \mathbf{1}_{\|S_N\| \ge 1}\right] \\ &\le \frac{B(B'+1)}{N^2}, \end{aligned}$$

where we use Markov's inequality to show that $\mathbb{P}\left(\|\bar{S}_N\| \ge 1\right) = \mathbb{P}\left(\|\bar{S}_N\|^3 \ge 1^3\right) \le B/N^2$. □

7 CONCLUSION AND DISCUSSION

In this paper, we revisited the rate of convergence of mean-field models. Most of the results in the literature show that the expected distance between the empirical measure $X^{(N)}$ and its mean-field approximation x converges to 0 at rate $O(1/\sqrt{N})$. Our approach is different : we study the distance between the expectation of $X^{(N)}$ and x and we show that this distance decreases as $O(1/N)$. The latter quantity is natural when one wants to compute expected performance indicators such as the average queue length or the marginal law of one object. We have shown that this convergence rate of $O(1/N)$ holds for finite and infinite-dimensional models. It holds for the transient regime and also for the stationary regime when the limiting ODE has a locally exponentially stable attractor.

We believe that the potential of applicability of our method is large. Most of the discrete-space mean-field models found in the literature satisfy our assumptions. For these models, our results provide new insights on how accurate mean-field approximation is. They also suggest that in many cases, the mean-field approximation can be refined by a term C/N. An easy way to evaluate this constant C is to compute the gap between the mean-field approximation and a simulation for a small system size ($N \approx 10$) and then to extrapolate to larger values of N. For future work, we are currently working on numerical and analytical methods to directly compute the constant C without simulation. This will provide new analytical and numerical tools to build approximations that depend on the system size N. We expect these new approximations to be more accurate than mean-field, especially when the system size is small (for example when N is a few tens).

8 ACKNOWLEDGMENT

The author would like to thank Romain Joly and Jaap Eldering for fruitful discussions on the link between asymptotic stability and the dependence on initial conditions. The author is grateful to Bruno Gaujal, Debankur Mukherjee and Sem Borst whose insightful comments substantially improved this paper. This work is partially supported by the EU project QUANTICOL, 600708.

REFERENCES

[1] F Baccelli, FI Karpelevich, M Ya Kelbert, AA Puhalskii, AN Rybko, and Yu M Suhov. 1992. A mean-field limit for a class of queueing networks. *Journal of statistical physics* 66, 3-4 (1992), 803–825.

[2] Michel Benaim and Jean-Yves Le Boudec. 2008. A class of mean field interaction models for computer and communication systems. *Performance Evaluation* 65, 11 (2008), 823–838.

[3] Luca Bortolussi and Richard A Hayden. 2013. Bounds on the deviation of discrete-time Markov chains from their mean-field model. *Performance Evaluation* 70, 10 (2013), 736–749.

[4] Henri Cartan. 1977. *Cours de calcul différentiel*. Hermann.

[5] F Cecchi, SC Borst, and JSH van Leeuwaarden. 2015. Mean-field analysis of ultra-dense csma networks. *ACM SIGMETRICS Performance Evaluation Review* 43, 2 (2015), 13–15.

[6] Augustin Chaintreau, Jean-Yves Le Boudec, and Nikodin Ristanovic. 2009. The Age of Gossip: Spatial Mean Field Regime. *SIGMETRICS Perform. Eval. Rev.* 37, 1 (June 2009), 109–120. DOI: http://dx.doi.org/10.1145/2492101.1555363

[7] Jeong-woo Cho, Jean-Yves Le Boudec, and Yuming Jiang. 2010. On the validity of the fixed point equation and decoupling assumption for analyzing the 802.11 mac protocol. *ACM SIGMETRICS Performance Evaluation Review* 38, 2 (2010), 36–38.

[8] Jaap Eldering. 2013. *Normally Hyperbolic Invariant Manifolds—the Noncompact Case (Atlantis Series in Dynamical Systems vol 2)*. Berlin: Springer.

[9] Patrick Eschenfeldt and David Gamarnik. 2016. Supermarket queueing system in the heavy traffic regime. Short queue dynamics. *arXiv preprint arXiv:1610.03522* (2016).

[10] Nicolas Gast and Gaujal Bruno. 2010. A Mean Field Model of Work Stealing in Large-scale Systems. *SIGMETRICS Perform. Eval. Rev.* 38, 1 (June 2010), 13–24. DOI: http://dx.doi.org/10.1145/1811099.1811042

[11] Nicolas Gast and Bruno Gaujal. 2012. Markov chains with discontinuous drifts have differential inclusion limits. *Performance Evaluation* 69, 12 (2012), 623–642.

[12] Nicolas Gast and Benny Van Houdt. 2015. Transient and steady-state regime of a family of list-based cache replacement algorithms. In *Proceedings of the 2015 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*. ACM, 123–136.

[13] Carl Graham. 2000. Chaoticity on path space for a queueing network with selection of the shortest queue among several. *Journal of Applied Probability* 37, 01 (2000), 198–211.
[14] Vassili N Kolokoltsov, Jiajie Li, and Wei Yang. 2011. Mean field games and nonlinear Markov processes. *arXiv preprint arXiv:1112.3744* (2011).
[15] Thomas G Kurtz. 1970. Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes. *Journal of Applied Probability* 7 (1970), 49–58.
[16] Thomas G Kurtz. 1978. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and Their Applications* 6, 3 (1978), 223–240.
[17] Yi Lu, Qiaomin Xie, Gabriel Kliot, Alan Geller, James R Larus, and Albert Greenberg. 2011. Join-Idle-Queue: A novel load balancing algorithm for dynamically scalable web services. *Performance Evaluation* 68, 11 (2011), 1056–1071.
[18] Malwina J Luczak and Colin McDiarmid. 2007. Asymptotic distributions and chaos for the supermarket model. *arXiv preprint arXiv:0712.2091* (2007).
[19] Wouter Minnebo and Benny Van Houdt. 2014. A fair comparison of pull and push strategies in large distributed networks. *IEEE/ACM Transactions on Networking (TON)* 22, 3 (2014), 996–1006.
[20] Michael David Mitzenmacher. 1996. *The Power of Two Random Choices in Randomized Load Balancing*. Ph.D. Dissertation. PhD thesis, Graduate Division of the University of California at Berkley.
[21] Charles Stein. 1986. Approximate computation of expectations. *Lecture Notes-Monograph Series* 7 (1986), i–164.
[22] John N Tsitsiklis and Kuang Xu. 2011. On the power of (even a little) centralization in distributed processing. *ACM SIGMETRICS Performance Evaluation Review* 39, 1 (2011), 121–132.
[23] Stephen RE Turner. 1998. The effect of increasing routing choice on resource pooling. *Probability in the Engineering and Informational Sciences* 12, 01 (1998), 109–124.
[24] Benny Van Houdt. 2013. A mean field model for a class of garbage collection algorithms in flash-based solid state drives. In *ACM SIGMETRICS Performance Evaluation Review*, Vol. 41. ACM, 191–202.
[25] Nikita Dmitrievna Vvedenskaya, Roland L'vovich Dobrushin, and Fridrikh Izrailevich Karpelevich. 1996. Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problemy Peredachi Informatsii* 32, 1 (1996), 20–34.
[26] Lei Ying. 2016. On the Approximation Error of Mean-Field Models. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*. ACM, 285–297.
[27] Lei Ying. 2016. On the Rate of Convergence of the Power-of-Two-Choices to its Mean-Field Limit. *CoRR* abs/1605.06581 (2016). http://arxiv.org/abs/1605.06581

Received October 2016; revised April 2017; accepted May 2017