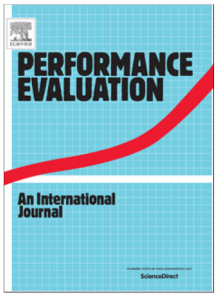
Accepted Manuscript



Size expansions of mean field approximation: Transient and steady-state analysis

Nicolas Gast, Luca Bortolussi, Mirco Tribastone

PII: S0166-5316(18)30263-3
DOI: https://doi.org/10.1016/j.peva.2018.09.005
Reference: PEVA 1970

To appear in: *Performance Evaluation*

Size Expansions of Mean Field Approximation: Transient and Steady-State Analysis

Nicolas Gast[a,*], Luca Bortolussi[b], Mirco Tribastone[c]

[a] *Inria, Univ. Grenoble Alpes, F-38000 Grenoble, France*
[b] *DMG, University of Trieste, Italy*
[c] *IMT School for Advanced Studies Lucca, Italy*

Abstract

Mean field approximation is a powerful tool to study the performance of large stochastic systems that is known to be exact as the system's size N goes to infinity. Recently, it has been shown that, when one wants to compute expected performance metric in steady-state, this approximation can be made more accurate by adding a term V/N to the original approximation. This is called a *refined* mean field approximation in [21].

In this paper, we improve this result in two directions. First, we show how to obtain the same result for the transient regime. Second, we provide a further refinement by expanding the term in $1/N^2$ (both for transient and steady-state regime). Our derivations are inspired by moment-closure approximation, a popular technique in theoretical biochemistry. We provide a number of examples that show: (1) that this new approximation is usable in practice for systems with up to a few tens of dimensions, and (2) that it accurately captures the transient and steady state behavior of such systems.

1. Introduction

Mean field approximation is a widely used technique in the performance evaluation community. The focus of this approximation is to study the performance of systems composed of a large number of interacting objects. Applications

*Corresponding author
Email address: nicolas.gast@inria.fr (Nicolas Gast)

range from biological models [46] to epidemic spreading [2] and computer-based systems [4]. In the performance evaluation community, this approximation has successfully been used to characterize the performance of CSMA protocols, [8], information spreading algorithms and peer-to-peer networks [9, 33], caching [10, 14, 20] or a quite popular subject such as load balancing strategies [18, 32, 36, 43, 45, 34, 48, 35]. This approximation can be used to study transient (for example the time to fill a cache [20]) or steady-state properties (for example the steady-state hit ratio [14, 10]).

One of the reasons of the success of mean field approximation is that it is often very accurate as soon as N, the number of objects in the system, exceeds a few hundreds. In fact, this approximation can be proven to be asymptotically exact as N goes to infinity, see for example [30, 31, 4, 19] and explicit bounds for the convergence rate exist [5, 15, 49, 51].

Recently, the authors of [21] proposed what they call a *refined* mean field approximation that can be used to characterize more precisely steady-state performance metrics. Their refinement uses that for many models, a steady-state expected performance metric of a system with N objects $\mathbb{E}[h(X)]$ is equal to its mean field approximation $h(\pi)$ plus a term in $1/N$:

$$\mathbb{E}[h(X)] = h(\pi) + \frac{1}{N}V_{(h)} + o\left(\frac{1}{N}\right), \tag{1}$$

where π is the fixed point of the ODE that describes the mean field approximation and $V_{(h)}$ is a constant that can be easily evaluated numerically.

By using a number of examples, they show that the refined approximation $h(\pi) + \frac{1}{N}V_{(h)}$ is much more accurate than the mean field approximation for moderate system sizes (i.e., a few tens of objects).

In this paper, we extend this method in two directions: First we generalize Equation (1) to the transient behavior; second we establish the existence of a second order term in $1/N^2$ (both in transient and steady-state regimes). More precisely, we establish conditions such that for any smooth function h, there exist constants $V_{(h)}$ and $A_{(h)}$ such that for any time $t \in [0;\infty) \cup \{\infty\}$:

$$\mathbb{E}\left[h(X(t))\right] = h(x(t)) + \frac{1}{N}V_{(h)}(t) + \frac{1}{N^2}A_{(h)}(t) + o\left(\frac{1}{N^2}\right). \tag{2}$$

We show that for the transient regime, $V_{(h)}(t)$ and $A_{(h)}(t)$ satisfy a linear time-inhomogeneous differential equation that can be easily integrated numerically (Theorem 1). The steady-state constants are directly computed from the fixed point of this linear differential equation (Theorem 3).

We use Equation (2) to propose two new approximations that depend on the system size N and that are expansions of the classical mean field approximation to the order $1/N$ and $1/N^2$, respectively. We then compare the following three approximations numerically on various examples:

- Mean field approximation: $h(x(t))$.
- $1/N$-expansion: $h(x(t)) + V_{(h)}(t)/N$.
- $1/N^2$-expansion: $h(x(t)) + V_{(h)}(t)/N + A_{(h)}(t)/N^2$.

Our numerical results shows that the two expansions capture very accurately the transient behavior of such a system even when $N \approx 10$. Moreover, they are generally much more accurate than the classical mean field approximation for small values of N (for transient and steady-state regimes). Our experiments also confirm that good accuracy of the $1/N$-expansion approximation that was observed for steady-state values in [21] : In most cases, the largest gain in accuracy comes from the $1/N$-term (both for the transient and steady-state values). The $1/N^2$ term does improve the accuracy but only marginally. We also study the limit of the method by studying an unstable mean field model that has an unstable fixed point. This last example has unique fixed point that is not an attractor which means that the classical mean field approximation cannot be used for steady-state approximation as shown in [4]. We show that in this case, the $1/N$ and $1/N^2$ expansions are not stable with time and are therefore inaccurate when the time becomes large.

To summarize, this paper makes theoretical contributions that are interesting from a practical perspective :

- Theoretical contributions – We show that the $1/N$-expansion proposed in [21] for steady-state estimation can be extended to the transient regime and can be refined to the next order correction term in $1/N^2$.
- Practical implications – We show that, despite the complexity of the formulas, it is relatively easy to compute the $1/N$ and $1/N^2$ terms (in the transient and steady-state regimes) for realistic models such as the supermarket model. The developed method is generic and is implemented in a tool [16].

Roadmap. The rest of the paper is organized as follows. We discuss related work in Section 2. We describe the model in Section 3. We develop the main results in Section 4 where we also provide the proofs. We show a simple malware propagation model in Section 5 in order to illustrate the main concepts. We then study the supermarket model in more detail in Section 6. In Section 7 we show an example that illustrates the limitations of the approach. Finally, we conclude in Section 8.

Reproducibility. The code to reproduce the paper – including simulations, figures and text – is available at `https://github.com/ngast/sizeExpansionMeanField` [17].

2. Related work

Our results apply to the classical density-dependent population process of Kurtz [31] of which the supermarket model of [45, 36] is an example.

2.1. Stein's Method

From a methodological point of view, our paper uses an approach similar to one of [15, 21, 29, 49, 51] in which the key idea is to compare an asymptotic expansion of the generator of the stochastic process with the generator of the mean field approximation, by using ideas inspired by Stein's method. In the papers [15, 29, 49, 51], this is used to obtain the rate of convergence of mean field models to their limit. In [21], this idea is used to compute the $1/N$-term for the steady-state behavior. The main theoretical contribution of the paper with respect to these is to show that this method can be pushed further to study transient regime and to obtain exact formula for the term in $1/N^2$. The work on Stein's method is not new [40] but has seen a regain of interest in the stochastic networks' community in the recent years thanks to the work of [6, 7].

2.2. System Size Expansion

Our paper is also closely related to an approach developed in the theoretical biology literature, known as system size expansion (SSE). The core idea of SSE dates back to the work of Van Kampen [44], and consists in working with the stochastic process expressing the fluctuations of the population model around the mean field limit, rescaled by $N^{-1/2}$, and approximating it by an absolute continuous process $\xi(t)$ taking real values. Starting from the Kolmogorov equation of the population model, and relying on a perturbation expansion, Van Kampen obtains an Fokker-Plank (FP) equation for $\xi(t)$ containing in the right hand side terms of order $N^{-p/2}$, for $p = 0, 1, \ldots$. Keeping only lower order terms (i.e. of order 0 and -1/2) results in a linear FP equation, whose solution is known as the *linear noise approximation*, which is equivalent to the central limit theorem proved by Kurtz [13].

Grima and coauthors, in [23] and following papers (see e.g. [25, 41, 42]), start from the FP and keep higher order terms of $\frac{1}{\sqrt{N}}$, introducing non-linear corrections to the linear approximation. The resulting FP cannot be solved exactly, but it can be used to derive differential equations for the mean, covariance, and potentially higher order moments. As far as the mean of the populations is concerned, the equation derived in [23, 25] shows an equivalent structure with the one obtained in this paper. The higher-order SSE equations, with corrections up to order N^{-1}, have been implemented in the tool iNA [41, 42], and more recently in the Matlab toolbox CERENA [27], the only working implementation to the authors' knowledge.

Even if equations for the mean population and for covariance of SSE and our method coincide, our approach has some advantanges. First of all, its derivation is rigorous and does not rely on any approximation of the process $\xi(t)$ being based on a perturbation expansion of the moment equations themselves. Secondly, it gives us an approximate equation for any function h of the population vector, which can be used to estimate higher order moments or hitting times. Finally, in this paper we validate our method with large-dimensional models : the $1/N$-expansion can be computed for models with hundreds of dimensions and the $1/N^2$-expansion can be computed for models with a few tens of dimensions.

2.3. Moment-closure Approximation

Our way of deriving the equations is also related to moment closure techniques [22], which work by truncating, at a finite order of moments, the exact infinite dimensional system of ODEs which captures the evolution in time of all moments of the population process. The truncation strategy typically assumes some form of the distribution, and uses the relationship among moments implied by that assumption to express high-order moments as a function of lower order ones (e.g. a Gaussian distribution has odd centered moments of order 3 and more all equal to zero). These techniques are in theory applicable to higher order moment — see for example [1] — but the approach presented in [1] seems difficult to apply in high dimensional models, due to the exponential dependence on the order of moments of the number of moment equations. The accuracy of moment closure approximations was studied in [24], and more recently in [38, 39]. These studies show that accuracy is subtle and hard to predict, and does not necessarily increase with the population size N. The method we present in this paper uses a more rigorous approach, rooted in convergence theorems, which guarantees exactness in the limit of large N, and can also be used to provide estimates of moments of any order without extra effort, by choosing proper functions h.

3. Model and notations

3.1. Density-Dependent Population Processes

We consider mean field models described by the classical model of density-dependent population process of [30]. A density dependent population process is a sequence of continuous time Markov chains $X^{(N)}$, where the index N is called the size of the system. For each N, the Markov chain $X^{(N)}$ evolves on a subset $\mathcal{E} \subset \mathbb{R}^d$, where d is called the *dimension* of the model. We assume that there exists a set of vectors $\mathcal{L} \in \mathcal{E}$ and a set of functions $\beta_{\boldsymbol{\ell}} : \mathcal{E} \to \mathbb{R}^+$ such that $X^{(N)}$ jumps from x to $x + \boldsymbol{\ell}/N$ at rate $N\beta_{\boldsymbol{\ell}}(x)$ for each $\boldsymbol{\ell} \in \mathcal{L}$.

Note that we state all our results using the framework of density-dependent population processes. An alternative would have been to used a continuous-time version of the discrete-time model of [4] for which our results can be adapted (see also the discussion in Section 2.3 of [21]).

3.2. Drift and Mean Field approximation

We define the drift f as

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

The drift is the expected variation of $X^{(N)}(t)$ when $X^{(N)}(t) = x$. By definition of the model, it is independent of N.

In all our results, we will assume that the ordinary differential equation (ODE) $\dot{x} = f(x)$ has a unique solution that starts in $x(0)$ at time 0 that we denote $t \mapsto \Phi_t x$. It satisfies: $\Phi_t x = x + \int_0^t \Phi_s x ds$. When it is not ambiguous, we will denote $x(t) := \Phi_t x$. The function $t \mapsto x(t)$ is called the *mean field approximation*.

3.3. Tensors, Derivatives and Einstein Notations

Our results rely on tensor computation. To simplify the expression of the results and their derivations, we use Einstein notation (also known as Einstein summation convention) that we recall here.

All vectors (or tensors) are d-dimensional (or of size $d \times d$, $d \times \cdots \times d$). For a given vector or tensor, the upper indices denote the component. For example, X^i denotes the ith component of a d-dimensional vector X, and C^{ijk} denotes the (i, j, k) components of a $d \times d \times d$-dimensional tensor C. We use the symbol $\otimes$ for the Kronecker product between two tensors: for two d-dimensional vectors

X and Y, $X \otimes Y$ denotes a $d \times d$-dimensional tensor whose component (i,j) is X^iY^j. Also, $Y^{\otimes 3} = Y \otimes Y \otimes Y$.

For a given function, the lower indices denote the variable on which we differentiate. Unless otherwise stated, the functions will always be evaluated at the mean field approximation $x(t)$. We use uppercase letters to denote the function evaluated at $x(t)$. To be more precise, this means that the quantity $F^i_{j_1\dots j_k}$ denotes the kth derivative of the ith component of f with respect to $x^{j_1}\dots x^{j_k}$ evaluated at $x(t)$:

$$F^i_{j_1\dots j_k} = \frac{\partial^k f^i}{\partial x^{j_1}\dots\partial x^{j_k}}(x(t))$$

We use Einstein summation convention, which implies summation over a set of repeated indices: each index variable that appears twice implies the summation over all the values of the index. For example $F^i_jV^j := \sum_j F^i_jV^j$ and $F^i_{j,k,\ell}B^{k,\ell} := \sum_{k,\ell} F^i_{j,k,\ell}B^{k,\ell}$. This convention greatly compactifies and therefore simplifies the expression of our results.

For a given $d^{\times k}$ tensor T, we denote by $\mathrm{Sym}(T)$ the symmetric part of a tensor, which is the summation of this tensor over all permutation of indices. Its $(i_1\dots i_k)$-component is:

$$\mathrm{Sym}(T)^{i_1\dots i_k} = \frac{1}{k!}\sum_{\sigma\in\mathfrak{S}_k} T^{i_{\sigma_1}\dots i_{\sigma_k}},$$

where $\mathfrak{S}_k$ is the symmetric group on k elements.

3.4. Summary of the Assumptions

In order to prove our results for the transient regime, we will use the following assumptions:

(A1) The sequence of stochastic processes $X^{(N)}$ is a density dependent process that evolves in a compact subset of $\mathcal{E} \subset \mathbb{R}^d$.

(A2) The drift function $f(x)$ is well defined and continuously differentiable four times. The function $q(x) = \sum_{\boldsymbol{\ell}\in\mathcal{L}} \boldsymbol{\ell}\otimes\boldsymbol{\ell}\beta_{\boldsymbol{\ell}}(x)$ is well defined and continuously differentiable twice. The function $r(x) = \sum_{\boldsymbol{\ell}\in\mathcal{L}} \boldsymbol{\ell}\otimes\boldsymbol{\ell}\otimes\boldsymbol{\ell}\beta_{\boldsymbol{\ell}}(x)$ is well defined and continuous.

Note that assumption (A2) on the differentiability of the drift, combined with assumption (A1) on the compactness of $\mathcal{E}$ implies that the drift is Lipschitz-continuous and bounded and that therefore the differential equation $\dot{x} = f(x)$ has a unique solution. These assumptions are mainly technical and are verified by many of the mean field models of the literature.

For the steady-state analysis, we will assume in addition:

(A3) For each N, the stochastic process $X^{(N)}$ has a unique stationary distribution.

(A4) The differential equation $\dot{x} = f(x)$ has a unique fixed point π that is a globally exponentially stable attractor, meaning that there exists two constants $a, b > 0$ such that for all $x \in \mathcal{E}$:

$$\|\Phi_t(x) - \pi\| \leq ae^{-bt}.$$

Assumption (A3) combined with the existence of a globally stable attractor is a natural condition when one wants to show that a stochastic model converges to the fixed point of its mean field approximation (this is often a necessary condition, as shown in [4, 11]). The exponential stability of this attractor is a natural condition to obtain rate of convergence for mean field models [49, 15]. Proving that a fixed point is an attractor is often difficult but showing that this attractor is exponentially stable is often much easier since it only depends on the eigenvalue properties of the Jacobian evaluated at the fixed point π.

4. Main results

In this section, we provide the main theoretical results. We start by stating the results for the transient case (§4.1), and the steady-state case (§4.2). We then comment on the numerical feasibility of the approach (§4.3) and we finish with the proofs (§4.4).

4.1. Transient Analysis

The main result of our analysis is Theorem 1, which characterizes how the moments of the difference between the stochastic system $X(t)$ and its mean field approximation evolve with time. We show that each of these moments admits an expansion with a first term in $1/N$ and a second term in $1/N^2$. The constants of this asymptotic expansion are characterized by a system of linear ODEs. One of the direct consequence of this theorem is Corollary 2 that provides an asymptotic expansion of the mean and the variance of $X^{(N)}$.

Theorem 1. *Under assumption (A1-A2), let $x(t)$ denote the unique solution of the ODE $\dot{x} = f(x)$ starting in $X^N(0)$. There exists a series of time-dependent tensors V, W, A, B, C and D such that, for any four time differentiable function $h : \mathbb{R}^d \to \mathbb{R}$, we have:*

$$\mathbb{E}\left[h(X^{(N)}(t))\right] = h(x(t)) + \frac{1}{N}\left(H_iV^i + \frac{1}{2}H_{ij}W^{ij}\right) \tag{3}$$
$$+ \frac{1}{N^2}\Big(H_iA^i + \frac{1}{2}H_{ij}B^{ij} + \frac{1}{6}H_{ijk}C^{ijk} + \frac{1}{24}H_{ijk\ell}D^{ijkl}\Big) + o\Big(\frac{1}{N^2}\Big),$$

where the terms $H_i \dots H_{ijk\ell}$ denotes the first to fourth derivative of h evaluated at $x(t)$.

The dimension of the tensors V and A is n; the dimension of W and B is $n \times n$; the dimension of C is $n \times n \times n$; the dimension of D is $n \times n \times n \times n$. For the $1/N$-terms, these tensors satisfy the following ODE system (with the initial conditions $V = 0$ and $W = 0$):

$$\dot{V}^i = F^i_jV^j + \frac{1}{2}F^i_{j,k}W^{j,k}$$
$$\dot{W}^{i,j} = F^i_kW^{k,j} + F^j_kW^{k,i} + Q^{i,j} = \mathrm{Sym}\Big(2F^i_kW^{kj}\Big) + Q^{ij}$$

For the $1/N^2$-terms, the ODE system is as follows (with the initial conditions $A = 0$,$B = 0$, $C = 0$ and $D = 0$)

$$\dot{A}^i = F^i_jA^j + \frac{1}{2}F^i_{j,k}B^{j,k} + \frac{1}{6}F^i_{j,k,\ell}C^{j,k,\ell} + \frac{1}{24}F^i_{j,k,\ell,m}D^{j,k,\ell,m}$$
$$\dot{B}^{ij} = \mathrm{Sym}\Big(2F^i_kB^{kj} + F^i_{k\ell}C^{k\ell j} + \frac{1}{3}F^i_{k\ell m}D^{k\ell mj}\Big)$$
$$+ Q^{ij}_kV^k + \frac{1}{2}Q^{ij}_{k\ell}W^{k\ell}$$
$$\dot{C}^{ijk} = \mathrm{Sym}\Big(3F^i_\ell C^{\ell jk} + \frac{3}{2}F^i_{\ell m}D^{\ell mjk} + 3Q^{ij}V^k + 3Q^{ij}_\ell W^{\ell k}\Big) + R^{ijk}$$
$$\dot{D}^{ijk\ell} = \mathrm{Sym}\Big(4F^i_mD^{mjk\ell} + 6Q^{ij}W^{k\ell}\Big).$$

where the symmetric $d \times d$ tensor Q and $d^{\times 3}$ tensor R are:

$$Q = \sum_{\boldsymbol{\ell} \in \mathcal{L}} (\boldsymbol{\ell} \otimes \boldsymbol{\ell})\beta_{\boldsymbol{\ell}}(x(t)) \tag{4}$$
$$R = \sum_{\boldsymbol{\ell} \in \mathcal{L}} (\boldsymbol{\ell} \otimes \boldsymbol{\ell} \otimes \boldsymbol{\ell})\beta_{\boldsymbol{\ell}}(x(t)); \tag{5}$$

The tensors Q_k and $Q_{k,\ell}$ correspond to the first and second derivatives of the function $x \mapsto \sum_{\boldsymbol{\ell}\in\mathcal{L}} \boldsymbol{\ell}\otimes\boldsymbol{\ell}\beta_{\boldsymbol{\ell}}(x)$, evaluated in $x(t)$:

$$Q_k = \frac{\partial}{\partial x^k}\sum_{\boldsymbol{\ell}\in\mathcal{L}}(\boldsymbol{\ell}\otimes\boldsymbol{\ell})\beta_{\boldsymbol{\ell}}(x(t))$$
$$Q_{k,\ell} = \frac{\partial^2}{\partial x^k\partial x^\ell}\sum_{\boldsymbol{\ell}\in\mathcal{L}}(\boldsymbol{\ell}\otimes\boldsymbol{\ell})\beta_{\boldsymbol{\ell}}(x(t)).$$

To prove this theorem, we will first prove the existence of the tensors and then will show that they satisfy the corresponding set of ODEs by computing how the moments evolve with time. In fact, an equivalent characterization of the tensors V, W,... is to use these tensors to construct asymptotic expansions of the moments of $X^{(N)}(t) - x(t)$. This is summarized in Corollary 2, which also has an interest in its own. This corollary also justifies why moment closure works: neglecting the first moment of $X^{(N)}(t) - x(t)$ gives the mean field approximation, neglecting the moment three and above gives the expansion of order $1/N$; finally neglecting the moments five and above gives the expansion of order $1/N^2$. In theory, it should be possible to continue the asymptotic expansion but the at the price of a much higher complexity in the expressions. In the numerical examples, we will show that the asymptotic development of the expectation provides a very accurate estimation of the true expectation in many cases.

Corollary 2. *Under the assumption of Theorem 1, we have*

$$\mathbb{E}\left[X^{(N)}(t) - x(t)\right] = \frac{1}{N}V(t) + \frac{1}{N^2}A(t) + o(1/N^2)$$
$$\mathbb{E}\left[(X^{(N)}(t) - x(t))^{\otimes 2}\right] = \frac{1}{N}W(t) + \frac{1}{N^2}B(t) + o(1/N^2)$$
$$\mathbb{E}\left[(X^{(N)}(t) - x(t))^{\otimes 3}\right] = \frac{1}{N^2}C(t) + o(1/N^2)$$
$$\mathbb{E}\left[(X^{(N)}(t) - x(t))^{\otimes 4}\right] = \frac{1}{N^2}D(t) + o(1/N^2)$$
$$\mathbb{E}\left[(X^{(N)}(t) - x(t))^{\otimes k}\right] = o(1/N^2) \qquad \text{for } k \geq 5.$$

In particular:

$$\operatorname{cov}(X^{(N)}(t), X^{(N)}(t)) = \frac{1}{N}W(t) + \frac{1}{N^2}(B(t) - V(t)\otimes V(t)) + o(1/N^2).$$

Proof. The first set of equation is a direct consequence of Theorem 1 applied to the functions $h(X) = (X - x)^{\otimes k}$ for $k = 1, 2 \ldots$.

For the covariance, we have :

$$\begin{aligned}\mathrm{cov}(X^{(N)}(t), X^{(N)}(t)) &= \mathbb{E}\left[(X^{(N)}(t) - x(t) + x(t) - \mathbb{E}\left[X^{(N)}(t)\right])^{\otimes 2}\right] \\ &= \mathbb{E}\left[(X^{(N)}(t) - x(t))^{\otimes 2}\right] - (x(t) - \mathbb{E}\left[X^{(N)}(t)\right])^{\otimes 2} \\ &= \frac{1}{N}W(t) + \frac{1}{N^2}(B(t) - V(t) \otimes V(t)) + o(1/N^2).\end{aligned}$$

□

4.2. Steady-State Regime

We now turn our attention to the steady-state regime. The next theorem shows that when the system in the mean field approximation has a unique attractor, then the tensors of Theorem 1 have a limit as t goes to infinity, and this limit can be used to obtain an asymptotic expansion in $1/N$ and $1/N^2$ in steady-state. For V and W, these equations are the same as ones developed in [21]. The novelty of this result is the $1/N^2$-expansion.

Theorem 3. *In addition to the assumption of Theorem 1, assume (A3) and (A3). Then the ODE of Theorem 1 also has a unique attractor. Moreover, in steady state for any four times differentiable function $h : \mathbb{R}^d \to \mathbb{R}$, one has:*

$$\begin{aligned}\mathbb{E}\left[h(X^{(N)})\right] &= h(\pi) + \frac{1}{N}\Big(H_i V^i + \frac{1}{2}H_{ij}W^{ij}\Big) \\ &+ \frac{1}{N^2}\Big(H_i A^i + \frac{1}{2}H_{ij}B^{ij} + \frac{1}{6}H_{ijk}C^{ijk} + \frac{1}{24}H_{ijk\ell}D^{ijkl}\Big) + o\Big(\frac{1}{N^2}\Big),\end{aligned}$$

where the terms $H_i \dots H_{ijk\ell}$ denotes the first to fourth derivative of h evaluated at the fixed point π and where the tensors satisfy the following system of linear equations:

$$2\mathrm{Sym}\Big(F^i_k W^{kj}\Big) = -Q^{ij} \qquad\qquad F^i_j V^j = \frac{1}{2}F^i_{jk}W^{jk}$$

and

$$\begin{aligned}&4\mathrm{Sym}(F^i_m D^{mjk\ell}) = -6\mathrm{Sym}(Q^{ij}W^{k\ell}) \\ &3\mathrm{Sym}(F^i_\ell C^{\ell jk}) = -\left(\mathrm{Sym}\left(\frac{3}{2}F^i_{\ell m}D^{\ell mjk} + 3Q^{ij}V^k + 3Q^{ij}_\ell W^{\ell k}\right) + R^{ijk}\right) \\ &2\mathrm{Sym}(F^i_k B^{kj}) = -\mathrm{Sym}\Big(F^i_{k\ell}C^{k\ell j} + \frac{1}{3}F^i_{k\ell m}D^{k\ell mj} + Q^{ij}_k V^k + \frac{1}{2}Q^{ij}_{k\ell}W^{k\ell}\Big) \\ &F^i_j A^j = -\Big(\frac{1}{2}F^i_{jk}B^{jk} + \frac{1}{6}F^i_{jk\ell}C^{jk\ell} + \frac{1}{24}F^i_{jk\ell m}D^{jk\ell m}\Big)\end{aligned}$$

where Q, R, Q_k and $Q_{k\ell}$ are evaluated at the fixed point π.

Also, as we will see in the proof, under the condition of Theorem 3, the convergence as N goes to infinity of Equation (3) is uniform in time. This is not necessarily the case when the mean field approximation does not have an attractor (see Section 7).

4.3. Computational Issues and Implementation

4.3.1. Transient Analysis

For a given mean field model, the ODE $\dot{x} = f(x)$ is an ODE of dimension d. As the drift f is in general non-linear, the solution $x(t)$ can rarely be computed in closed form but can be easily computed numerically for high dimensional models. Once the solution $x(t)$ is computed, the system of ODEs for V, W, A, B, C and D given by Theorem 1 is a system of linear ODEs with time-varying parameters.

The system of ODEs for V and W does not depend on A, B, C, D. It is therefore possible to compute the $1/N$ terms $V(t)$ and $W(t)$ by numerically integrating a system of $O(d^2)$ variables. The computation of the $1/N^2$ terms is more complicated because D has d^4 variables. This makes the computation of the $1/N^2$ terms feasible for d of at most a few tens.

4.3.2. Fixed-Point Analysis

The computation of the fixed point of Theorem 3 can also be solved by a numerical algorithm: The constants V to D are the solutions of a system of linear equations.

For the $1/N$-term, these equations are the same as the ones developed in [21] and can therefore be solved in $O(d^3)$ time in two steps:

- First, we obtain the matrix W from the solution of the Lyapunov equation $MW + (MW)^T = Q$ for some matrix M.
- Second, the vector V is the solution of a linear system of equations of dimension d.

The most costly step of the above is the computation of the solution of the Lyapunov equation, which can be done in $O(d^3)$ time by using the Bartels-Stewart algorithm [3].

Once the terms V and W have been computed, one can compute the tensors D, C, B, A (in this order) by exploiting the fact that the equation for D does not depend A, B, C (similarly, the equation for C does not depend on A and B; the equation for B does not depend on A). Each is a system of linear equations with respectively d^4, d^3, d^2 and d variables. For D and C, the system is a generalization of the classical Lyapunov equation $MW + (MW)^T = Q$ to higher order tensors. Although the system of linear equations is large, in our numerical examples we were able to solve these equations for systems as large as $d = 50$ dimensions in less than 20 seconds (which corresponds for D to a linear system with $50^4 = 6.25 \times 10^6$ unknowns).

4.3.3. Implementation

To compute numerically the mean field expansions, we implemented a generic tool in `Python` that can construct and solve the above equation. The tool is available at `https://github.com/ngast/rmf_tool/` [16]. It takes as an input a description of the model and uses symbolic differentiation to construct the derivatives of the drift and of the functions Q and R.

The tool uses the function `integrate.solve_ivp` of the library `scipy` [26] to numerically integrate the ODEs for computing $V(t)$ and $W(t)$ of Theorem 1. For the steady-state analysis, the tool uses the python library `scipy.sparse` to construct a sparse system of linear equations and the function `scipy.sparse.linalg.lgmres` to solve the sparse linear system.

Note that the use of symbolic differentiation makes the computation slow for large models. Hence, for the supermarket model, we directly implemented Python functions that compute the drift of the system and its derivative. All our specific implementation is available in the git repository of the paper [17].

4.3.4. Analysis of the computation time

To give a flavor of the numerical complexity of the method, we report in Figure 1 the time taken by our algorithm to compute the expansions for the supermarket model described in Section 6. This figure shows the computation times as a function of the number of dimensions of the model d. It contains four

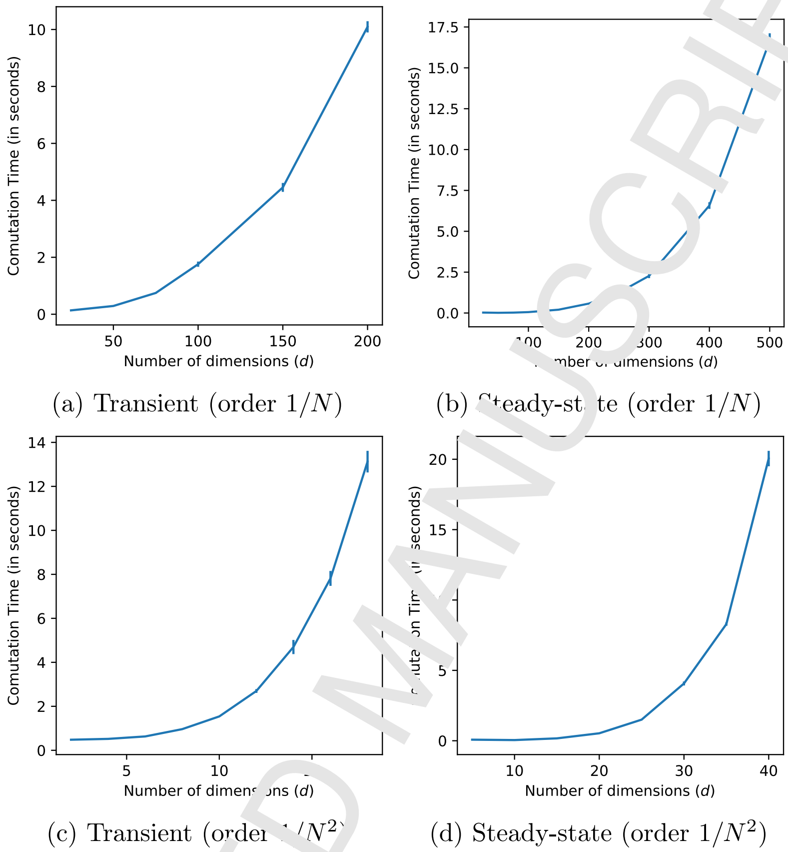


Figure 1: Supermarket : time to compute the approximation as a function of the number of dimension d. We compare the $1/N$-expansion (first line) and the $1/N^2$-expansions.

panels that correspond to:

(a) The time to compute $V(t)$ and $W(t)$ for $t \in [0, 10]$.

(b) The time to compute V and W of Theorem 3.

(c) The time to compute $A(t)$, $B(t)$, $C(t)$ and $D(t)$ for $t \in [0, 10]$.

(d) The time to compute A, B, C and D of Theorem 3.

We observe that, as expected, computing the time-varying constants of the transient regime is more costly than solving the fixed point equations because it requires solving an ODE: for a given time budget, one can compute the steady-state constants for a system of doubled size. Moreover, these results show that the computation of the $1/N$-terms $V(t)$ and $W(t)$ can be done for models with hundreds of dimensions in 10 seconds. With the same constraints of 10 seconds, the $1/N^2$-terms can be computed for models with a few tens of dimensions.

Note that we only provide this figure for the supermarket model because, among our three examples, it is the only one for which we can vary the dimension by changing the maximal queue lengths. We believe that the computation time does not grow too much with the dimension because the tensors corresponding to the derivatives of the drift or of the matrix Q are relatively sparse. The computation time might be higher for a model with denser tensors.

4.4. Proofs

To simplify the notation, where it is not needed in the proofs, we drop the superscript N and denote X instead of $X^{(N)}$.

4.4.1. Proof of Theorem 1

The proof of Theorem 1 is divided in two parts. We first we show the existence of the constants A, B,.. Second we show how to derive the ODE that they satisfy.

Existence of V, W – Here, we again use the notation $\Phi_s x$ to denote the value at time s of the solution of the ODE $\dot{x} = f(x)$ that starts in x at time 0. According to [15, Equation (19)] for any function $h : \mathcal{E} \to \mathbb{R}$, we have

$$\begin{aligned} N\,\mathbb{E}\left[h(\Phi_t x) - h(X^{(N)}(t)) \mid X_0^{(N)} = x\right] \\ = \int_0^t \mathbb{E}\left[\Delta^{(N)} h \circ \Phi_s(X^{(N)}(t-s)) \mid X_0^{(N)} = x\right] ds, \qquad (6) \end{aligned}$$

where $\Delta^{(N)}$ is the operator that, for a function g, gives the function $\Delta^{(N)} g$ defined by:

$$(\Delta^{(N)} g)(x) = N \sum_{\boldsymbol{\ell} \in \mathcal{L}} \beta_{\boldsymbol{\ell}}(x)(N(g(x + \frac{\boldsymbol{\ell}}{N}) - g(x)) - g_j(x)\boldsymbol{\ell}^j), \qquad (7)$$

where we recall the use of Einstein summation convention : $g_j(x)\boldsymbol{\ell}^j = \sum_{j=1}^d (\partial g(x))/(\partial x^j)\boldsymbol{\ell}^j$.

By using a Taylor expansion of g in the above equation, for a function $g:\mathbb{R}^n\to\mathbb{R}$ that is twice differentiable, we have

$$\Delta^{(N)}g(x)=\frac{1}{2}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\beta_{\boldsymbol{\ell}}(x)g_{ij}(x)\boldsymbol{\ell}^i\boldsymbol{\ell}^j+o(1/N) \tag{8}$$

where the hidden constant in the $o(1/N)$ depends on the modulus of continuity of the second derivative of g.

This shows that Equation (6) equals

$$\mathbb{E}\left[\int_0^t\frac{1}{2}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\beta_{\boldsymbol{\ell}}(X^{(N)}(t-s))(h\circ\Phi_s)_{ij}(X^{(N)}(t-s))\boldsymbol{\ell}^i\boldsymbol{\ell}^j ds\right]+o\left(\frac{1}{N}\right),$$

where $(h\circ\Phi_s)_{ij}(\Phi_{t-s}x)$ denotes the second derivative of $h\circ\Phi_s$ with respect to x^i and x^j evaluated at $\Phi_{t-s}x$. Again, the hidden constant in the $o(1/N)$ depends on the modulus of continuity of the second derivative of $(h\circ\Phi_s)$ which is finite for any time t because of Assumption (A2).

As $X^{(N)}(t-s)$ converges weakly to $\Phi_{t-s}x$ as N goes to infinity, the above quantity (to which Eq.(6) is equal) yields

$$\text{Eq.(6)}=\int_0^t\frac{1}{2}\sum_{\boldsymbol{\ell}\in\mathcal{L}}\beta_{\boldsymbol{\ell}}(\Phi_{t-s}x)(h\circ\Phi_s)_{ij}(\Phi_{t-s}x)\boldsymbol{\ell}^i\boldsymbol{\ell}^j ds+o(1/N). \tag{9}$$

In the quantity $(h\circ\Phi_s)_{ij}(\Phi_{t-s}x)$, the only dependence in h is a linear combination of the first and second derivative of h evaluated at $\Phi_t x$. Indeed, by the chain rule, for two functions g and h, the first and second derivative of $g\circ h$ evaluated in y is

$$(h\circ g)_i=(h_k\circ g)g_i^k$$
$$(h\circ g)_{ij}=(h_{k\ell}\circ g)g_i^k g_j^\ell+(h_k\circ g)g_{ij}^k$$

Replacing g by $\Phi_s x$ and evaluating the function is $\Phi_{t-s}x$ shows that the second derivative of $h\circ\Phi_s$ evaluated in $\Phi_{t-s}x$ is :

$$\begin{aligned}(h\circ\Phi_s)_{ij}(\Phi_{t-s}x)&=h_{k\ell}(\Phi_t x)(\Phi_s)_i^k(\Phi_{t-s}(x))(\Phi_s)_j^k(\Phi_{t-s}(x))\\&\quad+h_k(\Phi_t x)(\Phi_s)_{i,j}^k(\Phi_{t-s}(x)).\end{aligned}$$

Plugging this into Equation (9) shows that Equation (6) is equal to:

$$h_{k\ell}(\Phi_t x)\underbrace{\frac{1}{2}\int_0^t \sum_{\boldsymbol{\ell}\in\mathcal{L}} \beta_{\boldsymbol{\ell}}(\Phi_{t-s}x)(\Phi_s)_i^k(\Phi_{t-s}(x))(\Phi_s)_j^k(\Phi_{t-s}(x))\boldsymbol{\ell}^i\boldsymbol{\ell}^j ds}_{=:W^{k\ell}(t)}$$
$$+\, h_k(\Phi_t x)\underbrace{\frac{1}{2}\int_0^t \sum_{\boldsymbol{\ell}\in\mathcal{L}} \beta_{\boldsymbol{\ell}}(\Phi_s)_{i,j}^k(\Phi_{t-s}(x))\boldsymbol{\ell}^i\boldsymbol{\ell}^j ds}_{=:V^k(t)} + o(1/N).$$

This implies the existence of $V(t)$ and $W(t)$ in Equation (3).

Existence of $A\dots D$ – The proof of the existence of the terms A to D is similar. Hence, for space constraints we only sketch the main differences. The first ideas is to refined the expansion (8) to

$$\Delta^{(N)}g(x) = \frac{1}{2}\sum_{\boldsymbol{\ell}\in\mathcal{L}} \beta_{\boldsymbol{\ell}}(x)g_{ij}\boldsymbol{\ell}^i\boldsymbol{\ell}^j + \frac{1}{6N}\sum_{\boldsymbol{\ell}\in\mathcal{L}} \beta_{\boldsymbol{\ell}}(x)g_{ijk}\boldsymbol{\ell}^i\boldsymbol{\ell}^j\boldsymbol{\ell}^k + o(\frac{1}{N^2}). \tag{10}$$

This shows that Equation (6) equals

$$\mathbb{E}\left[\int_0^t \frac{1}{2}\sum_{\boldsymbol{\ell}\in\mathcal{L}} \beta_{\boldsymbol{\ell}}(X^{(N)}(t-s))(h\circ\Phi_s)_{ij}(X^{(N)}(t-s))\boldsymbol{\ell}^i\boldsymbol{\ell}^j ds\right]$$
$$+\mathbb{E}\left[\frac{1}{6N}\int_0^t\sum_{\boldsymbol{\ell}\in\mathcal{L}} \beta_{\boldsymbol{\ell}}(X^{(N)}(t-s))(h\circ\Phi_s)_{ijk}(X^{(N)}(t-s))\boldsymbol{\ell}^i\boldsymbol{\ell}^j\boldsymbol{\ell}^k ds\right] + o\Big(\frac{1}{N^2}\Big).$$

In the above equation, the second term is of order $1/N$ and involves the derivative up to order three of h. The first term is equal to (6) plus a correction term of order $1/N$ that involves the derivative up order four of h (evaluated at $\Phi_t x$).

Derivation of the ODEs – The evolution of the stochastic process $X(t) - x(t)$ can be decomposed in two parts : a jump part due to the fact that $X(t)$ jumps to $X(t) + \boldsymbol{\ell}/N$ at rate $N\beta_{\boldsymbol{\ell}}(X(t))$ and a drift part due to the fact $x(t)$ satisfies the ODE $\dot{x} = f(x)$. This shows that for any function h, one has :

$$\frac{d}{dt}\mathbb{E}\left[h(X(t)-x(t))\right]$$
$$= \sum_{\boldsymbol{\ell}\in\mathcal{L}} \mathbb{E}\left[\Big(h(X(t)-x(t)+\frac{\boldsymbol{\ell}}{N}) - h(X(t)-x(t))\Big)N\beta_{\boldsymbol{\ell}}(X(t))\right]$$
$$- \mathbb{E}\left[h_j(X(t)-x(t))f^j(x(t))\right].$$

In the above equation, the first line corresponds to the stochastic jumps of $X(t)$ while the second line corresponds the continuous variation of $x(t)$.

Applying the above equation[1] to the function $h(X) = (X - x)^{\otimes k}$ shows that :

$$
\begin{aligned}
&\frac{d}{dt}\mathbb{E}\left[(X-x)^{\otimes k}\right] && (11)\\
&\quad = \sum_{\boldsymbol{\ell}\in\mathcal{L}} \mathbb{E}\left[\left(\left(X - x + \frac{\boldsymbol{\ell}}{N}\right)^{\otimes k} - (X-x)^{\otimes k}\right) N\beta_{\boldsymbol{\ell}}(X)\right] \\
&\qquad\qquad - k\mathrm{Sym}\left(f(x)\otimes \mathbb{E}\left[(X-x)^{\otimes k-1}\right]\right) \\
&\quad = \sum_{m=1}^{k}\binom{k}{m}\mathrm{Sym}\left(\mathbb{E}\left[\frac{1}{N^{m-1}}\boldsymbol{\ell}^{\otimes m}\beta_{\boldsymbol{\ell}}(X)\otimes (X-x)^{\otimes k-m}\right]\right) \\
&\qquad - k\mathrm{Sym}\left(f(x)\otimes \mathbb{E}\left[(X-x)^{\otimes k-1}\right]\right) \\
&\quad = k\mathrm{Sym}\left(\mathbb{E}\left[(f(X)-f(x))\otimes (X-x)^{\otimes k-1}\right]\right) \\
&\qquad + \sum_{m=2}^{k}\binom{k}{m}\mathrm{Sym}\left(\mathbb{E}\left[\frac{1}{N^{m-1}}\boldsymbol{\ell}^{\otimes m}\beta_{\boldsymbol{\ell}}(X)\otimes (X-x)^{\otimes k-m}\right]\right) && (12)
\end{aligned}
$$

The the existence of the constants V, W, $A \dots D$ combined with Equation (12) show that the derivative of $\mathbb{E}[(X(t)-x(t))^{\otimes k}]$ admits an asymptotic expansion with a first term in $1/N$ and a second term in $1/N^2$. We are now ready to compute how the constants V, W, A... D evolve with time by computing the derivative with respect to time of $\mathbb{E}[(X-x)^{\otimes k}]$ for $k \in \{1\dots 4\}$ and identifying the $1/N$ and $1/N^2$ terms.

1. Case $\mathbb{E}[X-x]$ – By using Equation (12), we have :

$$\frac{d}{dt}\mathbb{E}\left[X-x\right] = \mathbb{E}\left[f(X)-f(x)\right].$$

Applying (3) to the function $h(X) = f^i(X) - f^i(x(t))$ implies that

$$
\begin{aligned}
\frac{d}{dt}\mathbb{E}\left[X^i - x^i\right] &= \frac{1}{N}(F^i_j V^j + \frac{1}{2}F^i_{jk}W^{jk}) \\
&+ \frac{1}{N^2}(F^i_j A^j + \frac{1}{2}F^i_{jk}B^{jk} + \frac{1}{6}F^i_{jkl}C^{jkl} + \frac{1}{24}F^i_{jklm}D^{jklm}) + o(\frac{1}{N^2})
\end{aligned}
$$

[1] In the remainder of the proof, we drop the dependence in t in most of the proof and write X instead of $X(t)$ and x instead of $x(t)$.

Using that $\frac{d}{dt}\mathbb{E}[X^i - x^i] = V^i/N + A^i/N^2 + o(1/N^2)$ and identifying the $O(1/N)$ and $O(1/N^2)$ terms shows that:

$$\frac{d}{dt}V^i = F^i_j V^j + \frac{1}{2}F^i_{jk}W^{jk}$$
$$\frac{d}{dt}A^i = F^i_j A^j + \frac{1}{2}F^i_{jk}B^{jk} + \frac{1}{6}F^i_{jk\ell}C^{jk\ell} + \frac{1}{24}F^i_{jk\ell m}D^{jk\ell m}$$

2. Case $\mathbb{E}[(X-x)^{\otimes 2}]$ – By using (12), we have

$$\frac{d}{dt}\mathbb{E}\left[(X-x)^{\otimes 2}\right] = 2\mathrm{Sym}(\mathbb{E}\left[(f(X)-f(x))\otimes(X-x)\right]) + \frac{1}{N}\mathbb{E}\left[q(X)\right],$$

where $q(X)$ is a covariance matrix defined by

$$q(X) = \sum_{\boldsymbol{\ell}\in\mathcal{L}} \beta_{\boldsymbol{\ell}}(X)\boldsymbol{\ell}\otimes\boldsymbol{\ell}.$$

For the first term, we consider the function $h(X) = ((f(X)-f(x))\otimes(X-x))^{ij}$ and we use Lemma 1(i). The first derivative of this function h evaluated at x is 0. The second derivative of h with respect to x^k and x^ℓ is $2\mathrm{Sym}(F_k \otimes J_{(\ell)})$, where $J_{(\ell)}$ is the matrix whose (ℓ,ℓ) element is one, the others being zero. The third derivative with respect to x^k, x^ℓ and x^m is equal to $3\mathrm{Sym}(F_{k\ell}\otimes J_{(m)})$. The fourth derivative with respect to x^k, x^ℓ, x^m and x^n is $4\mathrm{Sym}(F_{k\ell m}\otimes J_{(n)})$.

Hence, applying Equation (3) to h shows that

$$\frac{d}{dt}\mathbb{E}\left[\mathrm{Sym}((f(X)-f(x))\otimes(X-x))^{ij})\right]$$
$$= \mathrm{Sym}\Big(\frac{2}{2}F^i_k(\frac{1}{N}W^{kj} + \frac{1}{N^2}B^{kj}) + \frac{3}{6N^2}F^i_{k\ell}C^{k\ell j}$$
$$+ \frac{4}{24N^2}F^i_{k\ell m}D^{k\ell mj}\Big)$$

For the second term, applying (3) to the function q shows that

$$\mathbb{E}\left[q^{ij}(X)\right] = Q^{ij} + \frac{1}{N}Q^{ij}_k V^k + \frac{1}{2N}Q^{ij}_{k\ell}W^{k\ell} + O(1/N^2).$$

[1] Recall that the exponent ^{ij} stands for the component (ij).

This shows that:

$$\dot{W}^{ij} = 2\mathrm{Sym}(F^i_k W^{kj}) + Q^{ij}$$
$$\dot{B}^{ij} = \mathrm{Sym}\Big(2F^i_k B^{kj} + F^i_{k\ell} C^{k\ell j} + \frac{1}{3} F^i_{k\ell m} D^{k\ell mj} + Q^{ij}_k V^k + \frac{1}{2} Q^{ij}_{k\ell} W^{k\ell}\Big).$$

3. By using (12), with $\mathbb{E}[(X-x)^{\otimes 3}]$, we have :

$$\frac{d}{dt}\mathbb{E}\left[(X-x)^{\otimes 3}\right] = 3\mathrm{Sym}(\mathbb{E}\left[(f(X)-f(x)) \otimes (X-x)^{\otimes 2}\right]) + \frac{3}{N}\mathrm{Sym}(\mathbb{E}\left[q(X) \otimes (X-x)\right]) + \frac{1}{N^2}\mathbb{E}\left[r(X)\right], \quad (13)$$

where $r(x) = \sum_{\boldsymbol{\ell} \in \mathcal{L}} \boldsymbol{\ell}^{\otimes 3} \beta_{\boldsymbol{\ell}}(x)$.

To study the first term of Equation (13), we consider the function $h(X) = ((f(X)-f(x)) \otimes (X-x)^{\otimes 2})^{ijk}$. Applying Lemma 1(ii), the first two derivatives of this function evaluated at x are equal to 0. The third derivative of this function (with respect to x^ℓ, x^m, x^n) is equal to $6\mathrm{Sym}(F_\ell \otimes J_{(m)} \otimes J_{(n)})$ and the fourth derivative is equal to $12\mathrm{Sym}(F_{\ell m} \otimes J_{(m)} \otimes J_{(n)} \otimes J_{(o)})$.

Hence, applying Equation (3) to h shows that

$$\mathbb{E}\left[\mathrm{Sym}((f(X)-f(x)) \otimes (X-x)^{\otimes 2})\right] = \frac{1}{N^2}\mathrm{Sym}\left(\frac{6}{6} F^i_\ell C^{\ell jk} + \frac{12}{24} F^i_{\ell m} D^{\ell mjk}\right) + o(1/N^2).$$

The second term of Equation (13) can be treated by applying Equation (3) to $h(X) = q(X)(X^k - x^k)$, whose first derivative evaluated at x is Q^{ij} and whose second derivative is $2Q^{ij}_\ell \otimes J_{(\ell)}$ (see Lemma 1(i)). This shows that

$$\mathbb{E}\left[\mathrm{Sym}(q(X) \otimes (X-x)^{ijk})\right] = \frac{1}{N}\mathrm{Sym}\left(Q^{ij}V^k + \frac{2}{2} Q^{ij}_\ell W^{\ell k}\right) + o(\frac{1}{N}).$$

Finally, the last term of Equation (13) is equal to $R/N^2 + o(1/N^2)$.

This shows that Equation (13) has only terms in $O(1/N^2)$ plus term of order $o(1/N^2)$. By identifying the $O(1/N^2)$-terms, we get

$$\dot{C}^{ijk} = 3\mathrm{Sym}(F^i_\ell C^{\ell jk}) + \frac{3}{2}\mathrm{Sym}(F^i_{\ell m} D^{\ell mjk}) + 3\mathrm{Sym}(Q^{ij}V^k) + 3\mathrm{Sym}(Q^{ij}_\ell W^{\ell k}) + R^{ijk}$$

4. The derivative is similar for $\mathbb{E}[Y_t^{\otimes 4}]$. Applying (12) shows that

$$
\begin{aligned}
&\frac{d}{dt}\mathbb{E}\left[(X-x)^{\otimes 4}\right] \\
&= 4\mathrm{Sym}(\mathbb{E}\left[(f(X)-f(x))\otimes(X-x)^{\otimes 3}\right]) + \frac{6}{N}\mathbb{E}\left[q(X)\otimes(X-x)^{\otimes 2}\right] \\
&\quad + \frac{4}{N^2}\mathbb{E}\left[r(X)\otimes(X-x)^{\otimes 2}\right] + \frac{1}{N^3}\mathbb{E}\left[\sum_{\boldsymbol{\ell}\in\mathcal{L}}\beta_{\boldsymbol{\ell}}(X)\boldsymbol{\ell}^{\otimes 4}\right]
\end{aligned}
$$

By (3) with the function $h(x) = (f(X)-f(x))(X-x)^{\otimes 3}$ the first term is equal to $4\mathrm{Sym}(F_m^i D^{mjk\ell})/N^2 + o(1/N^2)$ (because the first three derivatives of this function h are equal to zero and the last one has a factor $4\times 3\times 2 = 24$ by Lemma 1(iii)).

For the second term, we can again use Equation (3) with $h(X) = q(X)(X-x)^2$ and Lemma 1(ii). The first derivative of h is zero and only the second term counts :

$$
\mathrm{Sym}(\mathbb{E}\left[q(X)\otimes(X-x)^{\otimes 2}\right]) = \frac{2}{2N}\mathrm{Sym}(Q\otimes W) + o(1/N).
$$

Finally, the one before last is of order $O(1/N^3)$ because of (3) and the last term is of order $O(1/N^3)$.

We therefore obtain :

$$
\dot{D}^{ijk\ell} = 4\mathrm{Sym}(F_m^i D^{mjk\ell}) + 6\mathrm{Sym}(Q^{k\ell}W^{ij}).
$$

In the above proof, we used the following lemma, whose proof is direct by using general Leibniz rule.

Lemma 1. *Let $g:\mathbb{R}\to\mathbb{R}$ be k-times differentiable. Then*

(i) $\frac{\partial^k(xg(x))}{(\partial x)^k} = xg^{(k)}(x) + kg^{(k-1)}(x)$

(ii) $\frac{\partial^k(x^2g(x))}{(\partial x)^k} = x^2g^{(k)}(x) + 2kxg^{(k-1)}(x) + k(k-1)g^{(k-2)}$

(iii) $\frac{\partial^k(x^3g(x))}{(\partial x)^k} = x^3g^{(k)}(x) + 3kx^2g^{(k-1)}(x) + 3k(k-1)xg^{(k-2)} + k(k-1)(k-2)g^{(k-3)}$

A.4.2. Proof of Theorem 3

Most of the work needed to prove Theorem 3 was already done in the proof of Theorem 1. Indeed, it should be clear the linear equations of Theorem 3 correspond to the fixed point equation of the ODE of Theorem 1. Therefore, to prove Theorem 3, the only remaining steps are to prove that:

1. These fixed point equations have a unique solution.
2. The system of ODEs of Theorem 1 converges to this solution.
3. One can exchange the limits $\lim_{t\to\infty}$ and $\lim_{N\to\infty}$

Uniqueness – the uniqueness of the solution, for V and W was already shown in [21]. For D, one can remark that its fixed point equation can be written as a matrix equation $M^{(4)}D = y$ where y is a vectorized version of $-6\mathrm{Sym}(Q\otimes W)$, and where the matrix $M^{(4)}$ is a $d^4\times d^4$ matrix that can be expressed as the Kronecker sum of four times the Jacobian of the drift evaluated at π:

$$M^{(4)}_{ijk\ell;abcd} = F^i_a\delta_{jb}\delta_{kc}\delta_{\ell d} + \delta_{ia}F^j_b\delta_{kc}\delta_{\ell d} + \delta_{ia}\delta_{jb}F^k_c\delta_{\ell d} + \delta_{ia}\delta_{jb}\delta_{kc}F^\ell_d, \tag{14}$$

where δ_{ij} is the Kronecker symbol that equals 1 if an only if $i=j$ and 0 otherwise. Note that in the above equation, the lines and columns of the matrix $M^{(4)}$ are indexed by the triples $ijk\ell$ (for the lines) or $abcd$ (for the columns).

By property the Kronecker sum, an eigenvalue of $M^{(4)}$ is the sum of four eigenvalues of the Jacobian matrix (F^i_j). As the system is exponentially stable, all the eigenvalues of the Jacobian matrix have negative real part. Therefore all eigenvalues of the matrix $M^{(4)}$ have negative real part and $M^{(4)}$ is invertible. This implies the existence and the uniqueness of the solution for D of the fixed point equation.

Once the D is fixed, the equation for C can be written is a similar way $M^{(3)}C = y$ where $M^{(3)}$ is the Kronecker sum of three times the Jacobian of the drift. A similar reasoning as the one for D shows that C is uniquely defined. This can be propagated to B and then A.

Convergence to the fixed point. The time-varying constant $D(t)$ satisfies a time inhomogeneous linear differential equation $\dot{D} = M^{(4)}(t)D + y(t)$, where $M^{(4)}(t)$ is the Kronecker sum of four times the Jacobian of the drift evaluated in

$x(t)$ and $y(t)$ (defined as in Equation (14)). As $x(t)$ converges to an exponentially stable attractor π, all eigenvalues of the Jacobian of the drift f evaluated in π have negative real part. This implies that there exists a time after which all eigenvalues of the Jacobian of f have negative real part in which case all eigenvalues of the matrix $M^{(4)}(t)$ have negative real part. This implies that the ODE for $D(t)$ is exponentially stable and that therefore $D(t)$ converges to the unique fixed point of this system. The same reasoning applies for C, B and A.

Exchange of the limits. The above steps guarantee that the terms $V(t)$ and $A(t)$ of the development in $1/N$ and $1/N^2$ converge as N goes to infinity. Informally, this shows that

$$\begin{aligned}\lim_{t\to\infty} \mathbb{E}\left[X^N(t)\right] - \pi &= \lim_{t\to\infty} \mathbb{E}\left[X^N(t) - x(t)\right] \\ &= \lim_{t\to\infty} \frac{1}{N}V(t) + \frac{1}{N^2}A(t) + o(1/N^2) \\ &= \frac{1}{N}V + \frac{1}{N^2}A + \lim_{t\to\infty} o(1/N^2). \end{aligned} \tag{15}$$

In order to conclude the proof, we need to show that it is possible to exchange the limits, which is to show that the term $\lim_{t\to\infty} o(1/N^2)$ is indeed a $o(1/N^2)$ term.

To see that, we use Stein's method and the ideas developed in [49, 15] to show that, in steady-state,

$$\mathbb{E}\left[h(X^{(N)})\right] - h(\pi) = \mathbb{E}\left[\Delta^{(N)} \int_0^\infty h(\Phi_s(X^{(N)}(s))) - h(\pi) ds\right],$$

where $\Delta^{(N)}$ is the operator defined in Equation (7). Note that this equation is a consequence of Equation (10) of [15] and is the analog of Equation (6) as t goes to infinity.

Concerning the exchangeability of the limits, for space constraints, we only sketch the main remaining ideas of the proofs. The first step is to show that the hidden constant of the $o(1/N^2)$ of Theorem 1 depends on the modulus of continuity of the function $G^{(t)}(x) = \int_0^t h(\Phi_s(x)) - h(\pi) ds$. This comes from Equation (10). The second idea is that the function $G(x) = \int_0^\infty h(\Phi_s(x)) - h(\pi) ds$ is four times differentiable and that the derivatives $G^{(t)}$ converge uniformly to the derivatives of G as t goes to infinity. This comes from perturbation theory : by [12, Lemma C.1], if the flow Φ has an exponentially stable attractor and is four times differentiable, then the first four derivatives of $\Phi_s(x)$ converge exponentially fast to 0. The same argument is used in the proof of Lemma 3.5 of [21]. These two arguments show that the modulus of continuity of the derivatives of $G^{(t)}$ are uniformly bounded in time and that therefore the convergence is uniform in time.

5. Example 1: Malware propagation

In this section we illustrate the above results with a simplified variant of the malware propagation model of [4, 28]. It can be viewed as an instance of a basic infection model in epidemiology (e.g. [37]). We choose this model because of its simplicity: since it is a one-dimensional model, the constants of the $1/N$ and the $1/N^2$ approximation can be computed in closed form and the stationary distribution can be evaluated numerically easily with high precision (it is a birth-death process). This allows us to assess the accuracy of the various approximations with high precision.

5.1. Model

We consider a model of malware propagation in a system composed of N agents. Each agent is either infected by the malware or not. Let X be the fraction of infected agents. We consider that each non-infected agent becomes infected at rate $1+X$ (the rate 1 corresponds to infection by an external source while the rate X corresponds an infection by a peer). An infected agent recovers at rate 1 due to some patching mechanism. This translates into the following transitions for X:

$$X \mapsto X + \frac{1}{N} \text{ at rate } N(1-X)(1+X)$$
$$X \mapsto X - \frac{1}{N} \text{ at rate } NX$$

5.2. Mean Field Approximations and Expansions

To apply Theorem 1 and 3, let us first compute the drift of the system, its derivative, the matrix Q and its derivative, and the tensor R. As the system is uni-dimensional, all tensors are in fact scalars. The drift is $f(x) = 1 - x^2 - x = r(x)$ and the function $q(x) = 1 - x^2 + x$. The ODE of the mean field approximation $\dot{x} = f(x)$ is a Bernoulli type equation, hence, the mean field approximation has the closed-form solution

$$x(t) = -\frac{1}{2} + \frac{\sqrt{5}}{2}\left(\frac{2}{1 - \alpha e^{-\sqrt{5}t}} - 1\right), \tag{16}$$

where $\alpha = (4x(0) + 1 - \sqrt{5})/(4x(0) + 1 + \sqrt{5})$ and $x(0)$ is the initial condition.

As there is a close form solution for the mean field approximation, it might be doable to obtain a close form expression for the constants $V(t)$, $W(t)$,... but the expressions of such constant seem highly complex. Hence, in our illustrations, we use our tool [16] to compute numerically these constants.

The fixed point analysis is simpler. From Equation (16), it is clear that the ODE $\dot{x} = f(x)$ has a unique attractor $\pi = (\sqrt{5} - 1)/2$ that is exponentially stable. Moreover, the derivatives of the drift (evaluated at π) are $f'(\pi) = -\sqrt{5}$, $f''(\pi) = 2$, $f^{(3)}(\pi) = f^{(4)}(\pi) = 0$. Finally, the function q evaluated at π is $q(\pi) = \sqrt{5} - 1$ and its derivatives are $q'(\pi) = 2 - \sqrt{5}$, $q''(\pi) = -2$. Last, we have that $r(\pi) = 0$.

After some algebra, it can be shown that the constants V and A that solve the fixed point equation of Theorem 3 are

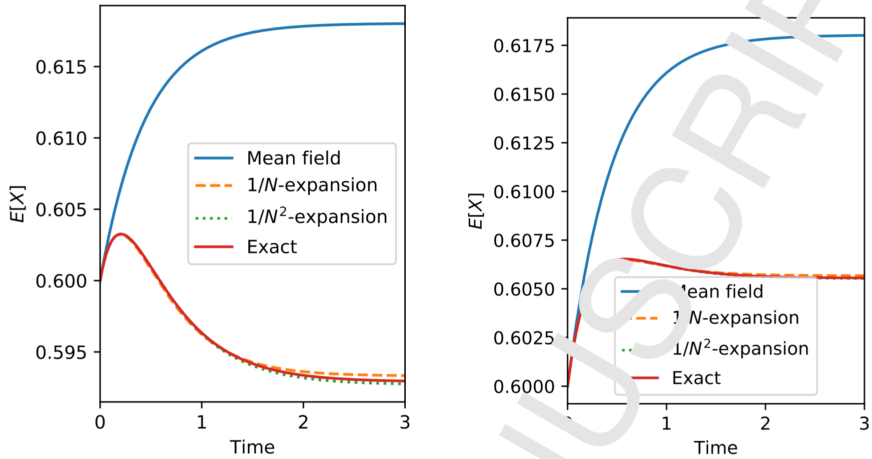
$$V = \frac{\sqrt{5} - 1}{10} \qquad \text{and} \qquad A = \frac{\sqrt{5} - 3}{50}.$$

Plugging the above quantity into Theorem 3 shows that, in steady-state and as N goes to infinity, one has :

$$\mathbb{E}[X] = \frac{\sqrt{5} - 1}{2}\left(1 - \frac{1}{5N}\right) + \frac{\sqrt{5} - 3}{50N^2} + o\left(\frac{1}{N^2}\right).$$

5.3. Numerical Comparison

In this section, we propose a numerical comparison of the exact values, the mean field approximation and the two expansions (up to order $1/N$ and $1/N^2$).



$N = 5$ $N = 10$

Figure 2: Malware model, transient regime: comparison of the mean field approximation, the $1/N$ and $1/N^2$ expansions and the exact value.

5.3.1. *Transient regime*

To perform a numerical comparison of the various approximations with the exact values, we implemented two numerical procedures. For the mean field approximation and the expansions, we implemented a numerical integration of the system of ODEs of Theorem 1. For the exact values, we used the fact that for a given size N, the stochastic model is a continuous time Markov chain with $N + 1$ states ($\{0, 1/N, 2/N, \dots, 1\}$). We again used a numerical integrator to integrate the Kolmogorov equations for this case.

The results are reported in Figure 2 in which we compare the three approximations (mean field and the two expansions) with the exact values, for $N = 5$ and $N = 10$. At the beginning, we start in a system where $X(0) = 0.6$ (*i.e.* $3N/5$ of the N agents are infected). We observe that the expansions provide a much better characterization of the transient regime that the classical mean field approximation. Note that for $N = 5$, the gain when going from the $1/N$ to the $1/N^2$ is small. For $N = 10$, the gain is almost invisible.

To observe more precisely what is the gain brought by the $1/N^2$ approximation, we plot in Figure 3 the $1/N^2$-constant $A(t)$ and compare it with the error

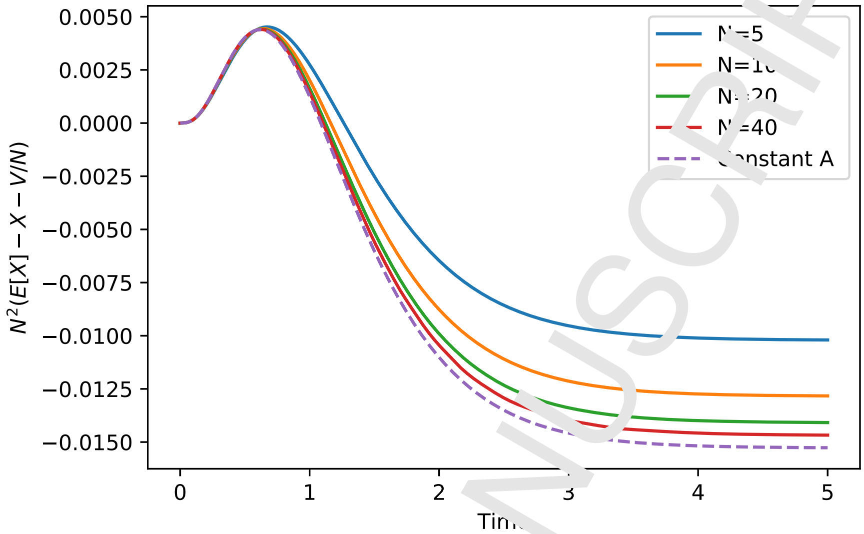


Figure 3: Malware model, transient regime : we compare the error of the expansion of order $1/N$ with the constant $A(t)$ of Theorem 1.

of the $1/N$-expansion rescaled by N^2: $N^2(\mathbb{E}[X(t)] - x(t) - V(t)/N)$, for various values of $N \in \{5, 10, 20, 30\}$. As shown by Theorem 1, the rescaled error of the $1/N$-expansion converges to $A(t)$ as N goes to infinity. This figures also shows that $A(t)$ is of order 10^{-2}. This explains why the gain in accuracy brought by the $1/N^2$-term is small: the error of the $1/N$-approximation is only around $0.01/N^2$.

5.3.2. Steady-state

We now verify the accuracy in steady-state. In Table 1, we verify the accuracy of the approximation for various values of $N \in \{1, 5, 10, 20, 30, 50\}$. We compare three values :

- $\mathbb{E}[X]$ that we computed by using the fact that this model is a birth-death process whose stationary measure can therefore be easily computed numerically.
- $\pi + V/N$, which is the refined approximation of [15] and that we call the $1/N$-expansion.

N	$\mathbb{E}[X]$	$1/N$-expansion		$1/N^2$-expansion	
			Error		Error
1	0.5000000	0.4944272	5.6e-03	0.4791486	2.1e-02
5	0.5929041	0.5933126	-4.1e-04	0.5927015	2.0e-04
10	0.6055449	0.6056733	-1.3e-04	0.6055205	2.4e-05
20	0.6118184	0.6118536	-3.5e-05	0.6118155	3.0e-06
30	0.6138977	0.6139138	-1.6e-05	0.6138968	8.7e-07
50	0.6155559	0.6155619	-5.9e-06	0.6155557	1.9e-07
∞	0.6180340	0.6180340	0	0.6180340	0

Table 1: Malware propagation model: comparison of the "true" expectation of X and the $1/N$ and $1/N^2$ expansions. The "error" column is the difference between $\mathbb{E}[X]$ and the expansion. Note that the classical mean field approximation is the value for $N = \infty$, which is $\pi \approx 0.6180340$.

- $\pi + V/N + A/N^2$ that we call the $1/N^2$-expansion.

We observe that for this model, the $1/N$ and $1/N^2$ expansions are already very accurate for $N = 1$ and they soon provide more than 4 digits of precision for $N \geq 10$. For $N \geq 10$, the error made by the $1/N^2$-expansion is an order of magnitude smaller than the error made by the $1/N$-expansion (the ratio between the two errors is approximately $0.6N$). The high accuracy of the $1/N$-expansion can be by the fact that the two constants are $V \approx 0.12$ and $A \approx -0.015$, hence, as for the transient regime, the difference between the two expansions is only $0.015/N^2$.

6. The supermarket model

We now focus on the classical supermarket model of [36, 45]. We study the gain of the $1/N$ and $1/N^2$ expansions for the transient and the steady-state regimes. As for the previous examples, the gain in accuracy of the $1/N$-expansion over the mean field approximation is large but the gain of the $1/N^2$-expansion over the $1/N$-expansion is smaller. Also, this model illustrates that it is possible to compute the $1/N$ and $1/N^2$ terms for a realistic model.

6.1. The Model

We consider a queuing system composed of N identical servers. Jobs arrive at a central broker according to a Poisson process of rate ρN and are dispatched towards the servers by using the JSQ(k) policy: for each incoming job, the broker samples k servers at random and sends the jobs to the server that has the smallest number of jobs in its queue (ties are broken at random). The time to process a job is exponentially distributed with mean 1.

This system can be modeled as a density dependent population process defined in Section 3. To see that, we assume that the queue size is bounded by d and we denote by $X_i(t)$ the fraction of servers with queue size i or more at time t. $X(t)$ is a Markov chain whose transitions are :

$$\begin{aligned} X &\to X - \tfrac{1}{N}\mathbf{e}_i \quad \text{at rate} \quad N(X_i - X_{i+1}) \\ X &\to X + \tfrac{1}{N}\mathbf{e}_i \quad \text{at rate} \quad N\rho(X_{i-1}^k - X_i^k), \end{aligned} \tag{17}$$

where $\mathbf{e}_i$ is a vector whose ith component is 1 the other ones being 0. Also, note that we use the classical notation for indices : X_i denotes the ith component of X and X_i^k denotes the kth power of X_i.

The explanation is as follows: A departure from a server with $i \geq 1$ jobs modifies X into $X - N^{-1}\mathbf{e}_i$ and occurs at rate $N(X_i - X_{i+1})$. An arrival at a server with i jobs modifies X into $X + N^{-1}\mathbf{e}_i$. Assuming that the k servers are picked with replacement, the least loaded among k servers has $i-1$ jobs with probability $X_{i-1}^k - X_i^k$.

6.2. Mean Field Approximation and Expansions

To apply Theorems 1 and 3, we first compute the drift, the constants Q, R and the needed derivatives.

The ith component of the drift of this model evaluated at x is F^i:

$$F^i = \rho(x_{i-1}^k - x_i^k) + (x_{i+1} - x_i). \tag{18}$$

The first derivative of the drift evaluated at a point x satisfies

$$F_{i-1}^i = k\rho x_{i-1}^{k-1}; \qquad F_i^i = -k\rho x_i^{k-1} - 1; \qquad F_{i+1}^i = 1,$$

all other terms being equal to 0.

Similarly, the second derivative satisfies

$$F_{i-1,i-1}^i = k(k-1)\rho x_{i-1}^{k-2} \qquad F_{ii}^i = -k(k-1)\rho x_i^{k-2},$$

all other terms being equal to 0. The expression is similar for the third and fourth derivatives.

The tensors Q and R of Equation (4) and (5) satisfy:

$$
\begin{aligned}
Q^{ii} &= (\beta_{e_i}(x) + \beta_{-e_i}(x)) = \rho(x_{i-1}^k - x_i^k) + (x_i - x_{i+1}) \\
R^{iii} &= F^i = \rho(x_{i-1}^k - x_i^k) + (x_{i+1} - x_i).
\end{aligned}
$$

Finally, the first and second derivatives of q evaluated in x satisfy

$$
\begin{aligned}
Q^{ii}_{i-1} &= k\rho x_{i-1}^{k-1} & Q^{ii}_{i} &= 1 - k\rho x_i^{k-1} & Q^{ii}_{i+1} &= -1 \\
Q^{ii}_{i-1,i-1} &= k(k-1)\rho x_{i-1}^{k-2} & Q^{ii}_{ii} &= -k(k-1)\rho x_i^{k-2}
\end{aligned}
$$

To apply Theorem 3, the only technical condition to verify is that the fixed point is exponentially stable. This is done for example in [49, 50]. The constants for the steady-state approximation can be computed by evaluating the above equation in π.

6.3. Algorithmic Considerations

In order to perform numerical comparison of the refined approximations and an estimation of the true values, we implemented various numerical algorithms. For the expected values, we implemented a C++ simulator of the supermarket model that simulates a density-dependent population process whose transitions are exactly the ones of Equation (17). For the transient analysis, to estimate the evolution of the expected queue length as a function of time, we performed an average of 10^5 (for $N = 10$) or 20000 (for $N = 20$) independent runs of simulations. This number of simulations is chosen as a compromise between computation time and accuracy. As we will observe in Figure 4, more simulations would give more accurate results but we choose to limit the computation time to $1h$ per panel. For the steady-state values, we compute the average of 1000 independent time-average of simulations after a warp-up period of $10000N$ events for each.

For the numerical analysis, we implemented a code to compute the parameters of the supermarket model and then use our tool [16] to solve numerically the ODEs of Theorem 1 or the fixed point equations of Theorem 3. As the size of the ODE for the $1/N^2$-approximation grows like d^4, we choose to bound the queue length to $d = 10$ for the $1/N^2$-expansions. In practice, using a larger maximal queue length brings to the same numerical value. For the transient regime, the computation time of the $1/N^2$-term is around 10sec and the one of the $1/N$-term less than one second. The computation of the fixed point is much faster than the one of the transient regime: it takes around 300ms for $d = 20$ and around 15s for $d = 50$ (on a 2013-laptop).

6.4. Numerical Comparisons

It is shown in [21] that the $1/N$-expansion provides estimates of the steady-state average queue length that are much more accurate than the classical mean field approximation. In this section we show that the $1/N$-expansion can also be used to improve the accuracy in the transient-regime and that the $1/N^2$-expansion improves on the $1/N$-expansion (both for transient and steady-state analysis).

6.4.1. Transient regime

We first consider how the expected queue length evolves with time. We consider the supermarket model with $k = 2$ choices and $\rho = 0.9$. We start in a system where the expected queue length is 2.8 : out of the N queues, $0.2N$ queues start with 2 jobs and $0.8N$ queues start with 3 jobs. We choose this value as it is close to 2.75, the steady-state average queue length predicted by the $1/N$-expansion for $N = 10$.

In Figure 4, we report how the expected queue length evolve with time compared to the three approximation (mean field, $1/N$-approximation and $1/N^2$-approximation). We observe in this figure that both for $N = 10$ and $N = 20$, the expansions provide an estimation of the evolution of the expected queue length that is much more accurate than the one provided by the classical mean field approximation. Moreover, for $N = 10$, the $1/N^2$-expansion provides a better approximation than the $1/N$-expansion. For $N = 20$, the two curves are almost indistinguishable.

N	k	ρ	Mean field	$1/N$-expansion	$1/N^2$-expansion	Simulation
10	2	0.7	1.1301	1.2150	1.2191	1.2193
20	2	0.7	1.1301	1.1726	1.1736	[illegible]
10	2	0.9	2.3527	2.7513	2.8045	[illegible]
20	2	0.9	2.3527	2.5520	2.5653	[illegible]
10	2	0.95	3.2139	4.1017	4.3265	[illegible]
20	2	0.95	3.2139	3.6578	3.7140	[illegible]
10	3	0.9	1.8251	2.2364	2.3322	[illegible]
20	3	0.9	1.8251	2.0307	2.0547	[illegible]
50	3	0.9	1.8251	1.9073	[illegible]	[illegible]
100	3	0.9	1.8251	1.8662	[illegible]	1.8672
10	4	0.95	2.0771	2.9834	[illegible]	3.3268
20	4	0.95	2.0771	2.5303	[illegible]	2.6376
50	4	0.95	2.0771	2.2584	[illegible]	2.2787
100	4	0.95	2.0771	2.1678	2.1766	2.1732

Table 2: Supermarket model, steady-state average queue length : comparison of the value computed by simulation with the three approximations.

For the simulation of the transient regime, the running time of simulation is approximately 0.1sec per run of our C++ simulator for $N = 20$ and 0.05sec for $N = 10$. This represents roughly 1h of computation for each of the two panels. As a comparison, the total time to compute the expansion of order $1/N^2$ is about 10 seconds (and does not depend on N), and the time to compute the expansion of order $1/N$ is around 1 second (using our python's implementation).

Note that we only present the results for $k = 2$ and $\rho = 0.9$. Similar results can be observed for other values of k and ρ with one difference: the smaller is ρ, the smaller is the difference between the approximations and the simulation (the difference between the $1/N$-expansion and the $1/N^2$-expansion can almost not be distinguished for $\rho < 0.7$). This is more visible in Table 2.

6.4.2. Steady-state

In Table 2, we present results that illustrate the accuracy of the expansions compared to the one of the classical mean field approximation. We choose a few values of k and ρ. More complete results can be found in the git repository of the paper [17].

We observe that in all tested cases, the $1/N$-expansion provides an estimation of the average queue length that is much more accurate than the one provided

	X_2	X_3	X_4	X_5	X_6	X_7
ρ=0.9, k=2, N=10						
Mean field	0.729	0.478	0.206	0.038	0.001	0.000
$1/N$-expansion	0.742	0.544	0.361	0.179	[illegible]	0.000
$1/N^2$-expansion	0.741	0.533	0.316	0.194	[illegible]	0.005
Simulation	0.741	0.534	0.327	0.170	0.077	0.032
ρ=0.95, k=2, N=20						
Mean field	0.857	0.698	0.463	0.204	0.039	0.001
$1/N$-expansion	0.861	0.721	0.544	0.371	0.184	0.026
$1/N^2$-expansion	0.861	0.719	0.527	[illegible]	0.210	0.122
Simulation	0.861	0.719	0.530	0.334	[illegible]	0.083
ρ=0.9, k=4, N=10						
Mean field	0.590	0.109	0.000	0.000	0.000	0.000
$1/N$-expansion	0.679	0.450	[illegible]	0.000	0.000	0.000
$1/N^2$-expansion	0.652	0.341	[illegible]	0.000	0.000	0.000
Simulation	0.657	0.344	0.140	0.051	0.018	0.006
ρ=0.95, k=4, N=20						
Mean field	0.774	0.341	[illegible]	0.000	0.000	0.000
$1/N$-expansion	0.802	[illegible]	[illegible]	0.000	0.000	0.000
$1/N^2$-expansion	0.795	0.429	0.578	0.001	0.000	0.000
Simulation	0.798	[illegible]	0.236	0.092	0.034	0.012

Table 3: Supermarket: steady-state distribution.

by the classical mean field approximation. The estimation provided by the $1/N^2$-expansion is generally more accurate but the gain brought by the $1/N^2$-term varies across the different parameters. The gain is the most visible for $k = 2$, in which case the $1/N^2$-expansion provides very accurate estimates, even for $N = 10$. This is less pronounced for $k = 3$ and $k = 4$, where the gain is more visible for higher values of N. Recall that in all cases, the mean field approximation provides estimates that do not depend on the system size N. They are systematically less accurate than the two expansions.

Theorem [illegible] can also be used to compute estimations of the queue length distribution. Indeed, for the supermarket model, $\mathbb{E}[X_i]$ is the probability that a given server has i jobs or more. In Table 3, we report the value of $\mathbb{E}[X_i]$ for various values of the parameters and $i \in \{2 \dots 7\}$. Note that we do not report the value $\mathbb{E}[X_1]$, which is the probability that a server is busy and is equal to ρ. We make two observations. First, for moderate values of ρ and k, the $1/N^2$-

expansion provides a very accurate estimation of the "true" distribution that we estimate by using simulation. This is less clear for higher values such as $d = 4$ and $\rho = 0.95$ for which the $1/N^2$ terms has a tendency to over correct for small values of N. Also, in all tested cases, the values for moderate values of i are well approximated, but the tail of the distribution is less well approximated. Note that for a fixed set of parameters (ρ, d), the two expansions become more accurate as N grows. This is illustrated in the git repository of the paper [17].

7. Limitations of the approach

In the previous examples, we concentrated on cases where the mean field approximation has a unique attractor, which implies that the mean field approximation and its expansions converge to the exact value of $\mathbb{E}[h(X)]$ uniformly in time (Theorem 3). In this section, we show that when the mean field approximation has a fixed point that is not an global attractor, this does not hold anymore. Moreover, in this setting, the two expansions do not work when t is too large compared to N.

7.1. An "Unstable" Malware Propagation Model

We consider a variation of the malware propagation example presented in Section 5 that is inspired by the model of [4]. The system is composed of N nodes. Each node can be dormant (D), active (A) or susceptible (S). Let X_D, X_A, X_S denote the proportion of dormant, active and susceptible nodes. A node that is dormant becomes active at rate $0.1 + 10X_A$. An active node becomes susceptible at rate 5 and a susceptible node becomes dormant at rate $1+\frac{10X_A}{X_D+\delta}$, where δ is a parameter of the model. This translates into the following transitions.

$$
\begin{aligned}
(X_D, X_A, X_S) &\mapsto (X_D - \frac{1}{N}, X_A + \frac{1}{N}, X_S) && \text{at rate } N(0.1 + 10X_A)X_D \\
(X_D, X_A, X_S) &\mapsto (X_D, X_A - \frac{1}{N}, X_S + \frac{1}{N}) && \text{at rate } N5X_A \\
(X_D, X_A, X_S) &\mapsto (X_D + \frac{1}{N}, X_A, X_S - \frac{1}{N}) && \text{at rate } N(1 + \frac{10X_A}{X_D + \delta})X_S
\end{aligned}
$$

This model satisfies all the assumptions (A1-A2) needed to apply Theorem 1 that characterize the transient regime. It also satisfies (A3): There exists a unique stationary distribution because for each system size N, the stochastic model is a finite state irreducible Markov chain. This model, however, does not satisfy assumption (A4) for all possible values of the parameter δ. Indeed, there exists a parameter value $\delta^* \approx 0.18$ such that the mean field limit has a unique attractor if and only if $\delta > \delta^*$. For $\delta < \delta^*$, the mean field approximation has a unique fixed point but unless the initial state is this fixed point, the limiting behavior of the solution of the ODE is an orbit. This is illustrated in Figure 5 where the two possible regimes are shown: for $\delta = 0.1$ the system has a stable orbit and an unstable fixed point. For $\delta = 0.5$ the system has a globally stable attractor.

It is known that when the mean field approximation has a globally stable attractor, then the sequence of stationary measures of the stochastic processes concentrates on this attractor as the system size N goes to infinity. On the other hand, when the mean field approximation has a (even unique) fixed point that is not an attractor (for example because there exist stable orbits), the sequence of stationary measures does not necessarily concentrate on this fixed point [4, 11].

When the stochastic model of size N is a finite-state irreducible continuous time Markov chain, it has a unique stationary distribution and $X^{(N)}(t)$ converges in distribution to a variable $X^{(N)}$ distributed according to this distribution. This shows that for any function h $\lim_{t\to\infty} \mathbb{E}[h(X^{(N)}(t))] = \mathbb{E}[h(X^{(N)})]$. Theorem 1 also shows that for any fixed time step t, $\lim_{N\to\infty} \mathbb{E}[h(X(t))] = h(x(t))$ where $x(t)$ is the mean field approximation. These reasons explain why one cannot exchange the limits $t \to \infty$ and $N \to \infty$:

$$\lim_{N\to\infty} \lim_{t\to\infty} \mathbb{E}\left[h(X^{(N)}(t))\right] \neq \lim_{t\to\infty} \lim_{N\to\infty} \mathbb{E}\left[h(X^{(N)}(t))\right] = \lim_{t\to\infty} h(x(t)),$$

because the limit on left hand side is independent of the initial condition of the Markov chain while the limit on the right-hand-side is not necessarily well defined if $x(t)$ does not converge to a unique fixed point regardless of the initial condition.

7.2. *Instability of the the Expansions*

One may hope that the expansions could be able to correct the non-exchangeability of the limits or at least would be able to compensate for some of the deviation. We show in fact in Figure 6 that not only the expansions do not correct the error of the mean field approximation but they can even make it worse when the mean field approximation has a limiting cycle (case $\delta = 0.1$).

To see that, we compare in Figure 6 the mean field approximation, the two expansions and an estimation of $\mathbb{E}[X(t)]$ obtained by simulation for the example described in Section 7.1 in the case where the fixed point is not an attractor ($\delta = 0.1$). We observe that for $N = 50$, the mean field approximation provides an accurate approximation of $\mathbb{E}[X(t)]$ for $t < 1$ and then starts oscillating for larger values of t whereas $\mathbb{E}[X(t)]$ stabilizes. The two expansions are slightly more accurate than the mean field approximation until $t \approx 1.2$. After this time, they diverge quickly and are much less accurate than the mean field approximation. The main explanation for this fact is that when the mean field approximation does not have an attractor, the ODE of Theorem 1 are unstable and the oscillations of the constants $V(t)$ and $A(t)$ grow with time. Note that the larger is N, the later the mean field approximation and its expansions start diverging from the expectation estimated by simulation.

When $\delta = 0.5$, the fixed point is an exponentially stable attractor. In this case, the error made by the mean field approximation (or by any of the two expansions) remains bounded with time, see Figure 7. Moreover in this case the expansions provide a more accurate estimate of the true value of $\mathbb{E}[X_A(t)]$. The behavior in this case is similar to the one observed for the two examples presented in the previous sections. Note that this examples is quite special in the sense that most of the mean field models studied in the queuing theory literature have a unique fixed point that is an attractor. This means that for these models it is more likely to observe a positive result like the one observed in Figure 7 rather than an oscillation like the one of Figure 6. This is no longer true when considering models from biochemistry [47].

8. Conclusion

In this paper, we show how mean field approximation can be refined by a term in $1/N$ and a second term $1/N^2$ where N is the size of the system. We exhibit conditions that ensure that this asymptotic expansion can be applied for the transient as well as the steady-state regimes. In the transient regime, these constants satisfy ordinary differential equations that can be easily integrated numerically. We provide a few examples that show that the $1/N$ and $1/N^2$ expansions are much more accurate than the classical mean field approximation. We also study the limitations of the approach and show that, when the mean field approximation does not have an attractor, these new approximations might be unstable for large time horizons. Obtaining a better approximation in this case remains a challenge that we leave for future work.

When we compare the accuracy of the classical mean field approximation to the one of the expansions of order $1/N$ and $1/N^2$, it seems that most of the gain in terms of accuracy are brought by the $1/N$-term. As the $1/N^2$-term is much more expensive to compute than the $1/N$ term, we believe that when the $1/N^2$-expansion is too hard to compute, staying with the $1/N$-expansion is already sufficient for many models. Finally, our derivation may also be exploited to obtain bounds on the error committed in the approximation of moments, which is something we aim at tackling as future work.

[1] Angelique Ale, Paul Kirk, and Michael PH Stumpf. 2013. A general moment expansion method for stochastic kinetic models. *The Journal of chemical physics* 138, 17 (2013), 174101.

[2] H. Andersson and T. Britton. 2000. *Stochastic Epidemic Models and Their Statistical Analysis.* Springer-Verlag.

[3] Richard H. Bartels and George W Stewart. 1972. Solution of the matrix equation AX+ XB= C [F4]. *Commun. ACM* 15, 9 (1972), 820–826.

[4] 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.

[5] 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.

[6] Anton Braverman and Jim Dai. 2017. Stein's method for steady-state diffusion approximations of $M/Ph/n+M$ systems. *The Annals of Applied Probability* 27, 1 (2017), 550–581.

[7] Anton Braverman, JG Dai, and Jiekun Feng. 2017. Stein's method for steady-state diffusion approximations: an introduction through the Erlang-A and Erlang-C models. *Stochastic Systems* 6, 2 (2017), 301–366.

[8] 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.

[9] 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. https://doi.org/10.1145/2492101.1555363

[10] Zhen Chen and LinZhang Lu. 2012. A projection method and Kronecker product preconditioner for solving Sylvester tensor equations. *Science China Mathematics* 55, 6 (2012), 1281–1292.

[11] Jeongwoo Cho, Jean-Yves Le Boudec, and Yuming Jiang. 2012. On the asymptotic validity of the decoupling assumption for analyzing 802.11 MAC protocol. *IEEE Transactions on Information Theory* 58, 11 (2012), 6879–6893.

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

[13] Stewart N. Ethier and Thomas G. Kurtz. 2009. *Markov processes: characterization and convergence*. Vol. 282. Wiley.

[14] Ronald Fagin. 1977. Asymptotic miss ratios over independent references. *J. Comput. System Sci.* 14, 2 (1977), 222–250.

[15] Nicolas Gast. 2017. Expected values estimated via mean-field approximation are 1/N-accurate. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 1, 1 (2017), 17.

[16] Nicolas Gast. 2018. Refined Mean Field Tool. https://github.com/ngast/rmf_tool.

[17] Nicolas Gast, Luca Bortolussi, and Mirco Tribastone. 2018. Size Expansions of Mean Field Approximation: Transient and Steady-State Analysis. https://github.com/ngast/sizeExpansionMeanField.

[18] 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. https://doi.org/10.1145/1811099.1811042

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

[20] 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.

[21] Nicolas Gast and Benny Van Houdt. 2017. A Refined Mean Field Approximation. *Proc. ACM Meas. Anal. Comput. Syst* 1 (2017).

[22] Colin S Gillespie. 2009. Moment-closure approximations for mass-action models. *IET systems biology* 3, 1 (2009), 52–58.

[23] Ramon Grima. 2010. An effective rate equation approach to reaction kinetics in small volumes: Theory and application to biochemical reactions in nonequilibrium steady-state conditions. *The Journal of chemical physics* 133, 3 (2010), 07B604.

[24] Ramon Grima. 2012. A study of the accuracy of moment-closure approximations for stochastic chemical kinetics. *The Journal of chemical physics* 136, 15 (2012), 04B616.

[25] Ramon Grima, Philipp Thomas, and Arthur V. Straube. 2011. How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations? *The Journal of Chemical Physics* 135, 8 (2011).

[26] Eric Jones, Travis Oliphant, Pearu Peterson, et al. 2001–. SciPy: Open source scientific tools for Python. http://www.scipy.org/ [Online; accessed 2017-06-15].

[27] Atefeh Kazeroonian, Fabian Fröhlich, Andreas Raue, Fabian J. Theis, and Jan Hasenauer. 2016. CERENA: ChEmical REaction Network Analyzer?A Toolbox for the Simulation and Analysis of Stochastic Chemical Kinetics. *PloS one* 11, 1 (2016), e0146732.

[28] MHR Khouzani, Saswati Sarkar, and Eitan Altman. 2012. Maximum damage malware attack in mobile wireless networks. *IEEE/ACM Transactions on Networking* 20, 5 (2012), 1347–1360.

[29] Vassili N Kolokoltsov, Jiajie Li, and Wei Yang. 2011. Mean field games and nonlinear Markov processes. *arXiv preprint arXiv:1112.3744* (2011).

[30] Thomas G Kurtz. 1970. Solutions of Ordinary Differential Equations as Limits of Pure Jump Markov Processes. *Journal of Applied Probability* 7 (1970), 49–58.

[31] Thomas G Kurtz. 1978. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and Their Applications* 6, 3 (1978), 223–240.

[32] 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.

[33] L. Massoulié and M. Vojnović. 2005. Coupon Replication Systems. *SIGMETRICS Perform. Eval. Rev.* 33, 1 (June 2005), 2–13. https://doi.org/10.1145/1071690.1064215

[34] W. Minnebo and B. Van Houdt. 2014. A Fair Comparison of Pull and Push Strategies in Large Distributed Networks. *IEEE/ACM Transactions on Networking* 22 (2014), 996–1006. Issue 3.

[35] Michael Mitzenmacher. 2016. Analyzing distributed Join-Idle-Queue: A fluid limit approach. In *Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on*. IEEE, 312–318.

[36] 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.

[37] James D. Murray. 2002. *Mathematical Biology I: An Introduction* (3rd ed.). Springer.

[38] David Schnoerr, Guido Sanguinetti, and Ramon Grima. 2014. Validity conditions for moment closure approximations in stochastic chemical kinetics. *The Journal of Chemical Physics* 141, 8 (2014), 084103. https://doi.org/10.1063/1.4892838

[39] David Schnoerr, Guido Sanguinetti, and Ramon Grima. 2015. Comparison of different moment-closure approximations for stochastic chemical kinetics. *The Journal of Chemical Physics* 143, 18 (2015), 185101. https://doi.org/10.1063/1.4934990

[40] Charles Stein. 1986. Approximate computation of expectations. *Lecture Notes-Monograph Series* 7 (1986), i–164.

[41] Philipp Thomas, Hannes Matuschek, and Ramon Grima. 2012. Intrinsic Noise Analyzer: A Software Package for the Exploration of Stochastic Biochemical Kinetics Using the System Size Expansion. *PLoS ONE* 7, 6 (2012), e38518. `https://doi.org/10.1371/journal.pone.0038518`

[42] Philipp Thomas, Hannes Matuschek, and Ramon Grima. 2013. Computation of biochemical pathway ?uctuations beyond the linear noise approximation using iNA. (2013), 5.

[43] 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.

[44] N. G. van Kampen. 2007. *Stochastic processes in physics and chemistry.* Elsevier, Amsterdam; Boston; London.

[45] 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.

[46] D. ~J Wilkinson. 2006. *Stochastic Modelling for Systems Biology.* Chapman & Hall.

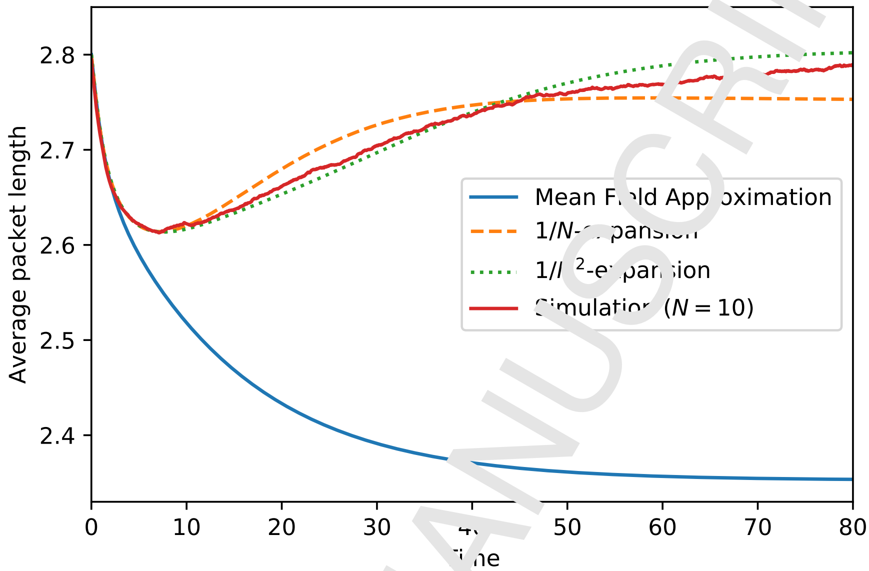
[47] D. J Wilkinson. 2011. *Stochastic Modelling for Systems Biology, 2nd edition.* CRC Press, Inc.

[48] Q. Xie, X. Dong, Y. Lu, and R. Srikant. 2015. Power of D Choices for Large-Scale Bin Packing: A Loss Model. *SIGMETRICS Perform. Eval. Rev.* 43, 1 (June 2015), 321–334. `https://doi.org/10.1145/2796314.2745849`

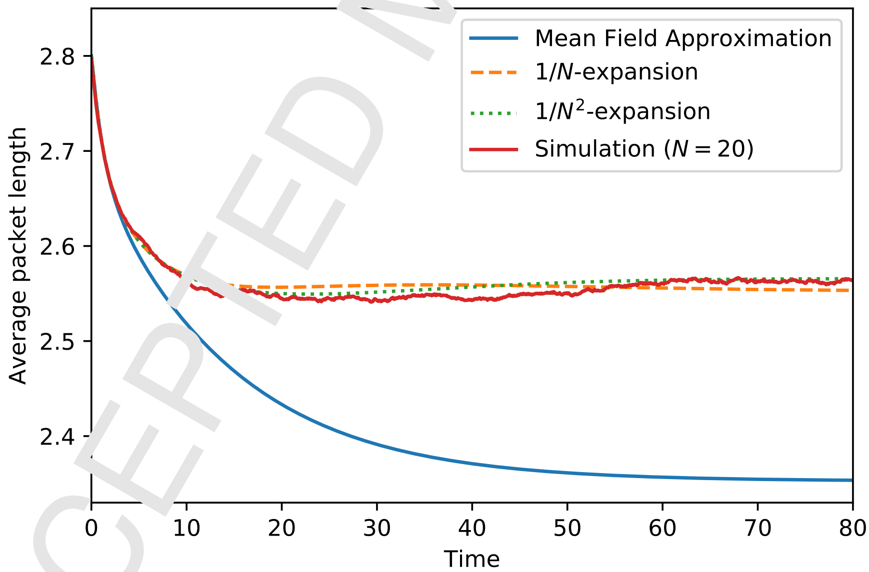
[49] Lei Ying. 2016. On the Approximation Error of Mean-Field Models. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science.* ACM, 285–297.

[50] 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`

[51] Lei Ying. 2017. Stein's Method for Mean Field Approximations in Light and Heavy Traffic Regimes. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 1, 1 (2017), 12.

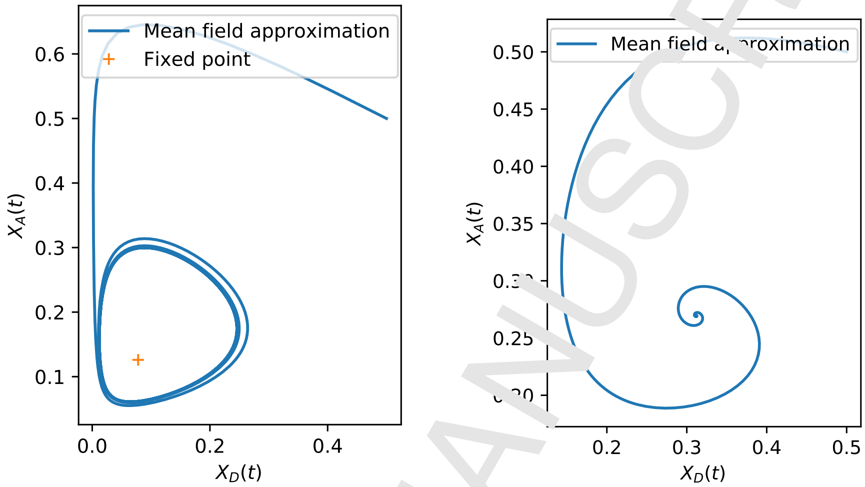


$N = 10$, $\rho = 0.9$, $k = 2$.



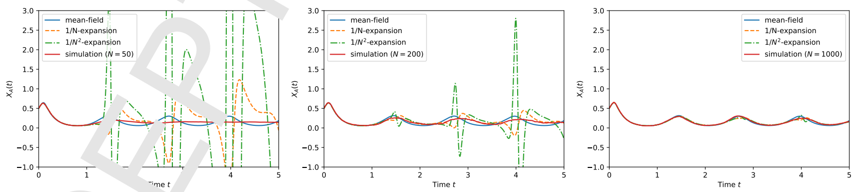
$N = 20$, $\rho = 0.9$, $k = 2$.

Figure 4: Supermarket model and transient regime: Comparison of the classical mean field approximation and the two expansions with data from simulations.



Fixed point $\neq$ attractor ($\delta = 0.1$) Fixed point $=$ attractor ($\delta = 0.5$)

Figure 5: The unstable malware model: illustration of the two possible regimes of the mean field approximation.



(a) $N = 50$ (b) $N = 200$ (c) $N = 1000$

Figure 6: "Unstable" malware model : when the fixed point is not an attractor ($\delta = 0.1$), the accuracy of the approximations is not uniform in time for a fixed system size N.

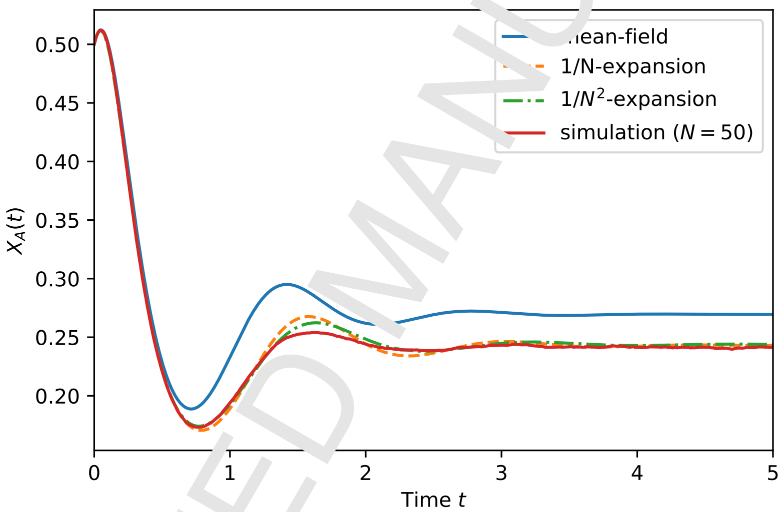


Figure 7: "Unstable" malware model in the stable case $\delta = 0.5$ and $N = 50$ (complement of Figure 6).