The averaging principle

Gadi Fibich [*], **Arieh Gavious** [†], **Eilon Solan** [*]

[*]Tel Aviv University, Israel, and [†]Ono Academic College and Ben Gurion University, Israel

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Typically, models with a heterogeneous property are considerably harder to analyze than the corresponding homogeneous models, in which the heterogeneous property is replaced with its average value. In this study we show that any outcome of a heterogeneous model that satisfies the two properties of *differentiability* and *interchangibility*, is $O(\epsilon^2)$ equivalent to the outcome of the corresponding homogeneous model, where ϵ is the level of heterogeneity. We then use this *averaging principle* to obtain new results in queueing theory, game theory (auctions), and social networks (marketing).

mathematical modeling | heterogeneity | averaging | homogenization

Mathematical modeling is a powerful tool in scientific research. Typically, the mathematical model is merely an approximation of the actual problem. Therefore, when choosing the model to work with, one has to strike a balance between complex models that are more realistic, and simpler models that are more amenable to analysis and simulations. This dilemma arises, for example, when the model contains a heterogeneous quantity. In such cases, a huge simplification is usually achieved by replacing the heterogeneous quantity with its average value. The natural question that arises is whether this approximation is "legitimate", i.e., whether the error that is introduced by this approximation is sufficiently small.

Let us illustrate this with the following example, which will be discussed in details later on. Consider a queue with k heterogeneous servers, whose expected service times are $\{\mu_1, \ldots, \mu_k\}$. We would like to calculate analytically the expected number of customers in the system, which we denote by $F(\mu_1, \ldots, \mu_k)$.[1] While an explicit expression for $F(\mu_1, \ldots, \mu_k)$ is not available, there is a well-known explicit expression in the case of k homogeneous servers, which we denote by $F_{\text{homog.}}(\mu) := F(\underbrace{\mu, \ldots, \mu}_{\times k})$. A natural approximation for the expected number of customers in the system is

$$F(\mu_1, \ldots, \mu_k) \approx F_{\text{homog.}}(\bar{\mu}), \qquad [1]$$

where $\bar{\mu}$ is the average of $\{\mu_1, \ldots, \mu_k\}$.

More generally, let $F(\mu_1, \ldots, \mu_k)$ denote an "outcome" of a heterogeneous model, let

$$\epsilon = \frac{\max_{1 \le i \le k} |\mu_i - \bar{\mu}|}{|\bar{\mu}|}, \qquad [2]$$

denote the level of heterogeneity of $\{\mu_1, \ldots, \mu_k\}$, and let $F_{\text{homog.}}(\mu)$ denote the outcome of the corresponding homogeneous model. If the function $F(\mu_1, \ldots, \mu_k)$ is differentiable, then it immediately follows that

$$F(\mu_1, \ldots, \mu_k) = F_{\text{homog.}}(\bar{\mu}) + O(\epsilon).$$

Therefore, roughly speaking, for a 10% heterogeneity level, the error of approximating $F(\mu_1, \ldots, \mu_k)$ with $F_{\text{homog.}}(\bar{\mu})$ is O(10%). In many studies in different fields, however, researchers have noted that the error of this approximation is considerably smaller than $O(\varepsilon)$. Moreover, this observation seems to hold even when the level of heterogeneity is not small.

In this study we show that these observations follow from a general principle, which we call the *Averaging Principle*. Specifically, we show that any outcome of a heterogeneous model that satisfies the two properties of *differentiability* and *interchangeability*, is $O(\epsilon^2)$ asymptotically equivalent to the outcome of the corresponding homogeneous model, i.e.,

$$F(\mu_1, \ldots, \mu_k) = F_{\text{homog.}}(\bar{\mu}) + O(\epsilon^2).$$

Thus, if the function F is also interchangeable, the error of the approximation **[1]** for a 10% heterogeneity level is only O(1%).

The Averaging Principle

Let $F(\mu_1, \ldots, \mu_k)$ be an outcome of a model with a heterogeneous property, captured by the k parameters $\{\mu_1, \ldots, \mu_k\}$, that satisfies the following two properties:

1. **Differentiability**: F is twice-differentiable at and near the diagonal $\mu_1 = \cdots = \mu_k$.
2. **Interchangeability**: For every $(\mu_1, \ldots, \mu_k) \in \mathbb{R}^k$ and every $i \neq j$, $F(\ldots, \mu_i, \ldots, \mu_j, \ldots) = F(\ldots, \mu_j, \ldots, \mu_i, \ldots)$. Thus, the outcome F is independent of the identities/indices of the heterogeneous parameters.[2]

Then, we have the following result:[3]

Theorem 1. (The Averaging Principle) *Let F satisfy the differentiability and interchangeability properties. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)$ be "sufficiently close to the diagonal", i.e.,*

$$||\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}_A|| < C_{\bar{\mu}_A},$$

where $\bar{\boldsymbol{\mu}}_A = (\underbrace{\bar{\mu}_A, \ldots, \bar{\mu}_A}_{\times k})$, $\bar{\mu}_A = \frac{1}{k}\sum_{j=1}^{k} \mu_j$ is the arithmetic average, $||\cdot||$ is a vector norm on $\mathbb{R}^k$, and $C_{\bar{\mu}_A}$ is a positive constant that only depends on $\bar{\mu}_A$ (and of course on F). Then,

$$F(\mu_1, \ldots, \mu_k) = F_{\text{homog.}}(\bar{\mu}_A) + O(||\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}_A||^2), \qquad [3]$$

where $F_{\text{homog.}}(\mu) := F(\underbrace{\mu, \ldots, \mu}_{\times k})$.

Theorem 1 remains valid if $\{\mu_1, \ldots, \mu_k\}$ are functions and not scalars, see the Game Theory (auctions) example below.

Reserved for Publication Footnotes

[1] In order to focus on the heterogeneous property, we suppress the dependence of F on other parameters.

[2] For example, in the queueing-system example, switching the identities/locations of two servers does not affect the expected number of customers in the system.

[3] This and all other proofs are given in the Appendix.

Using the geometric and harmonic averages. In Theorem 1 we averaged the $\{\mu_i\}$s using the arithmetic mean. It is well-known in homogenization theory that in some cases the correct homogenization is provided by the geometric or the harmonic mean. To address the question of the "correct" averaging, we recall the following result:

Lemma 1. *Let $\mu > 0$, and let $\{h_1, \ldots, h_k\} \in \mathbb{R}$. Then, as $\epsilon \to 0$, the arithmetic, geometric and harmonic means of $\{\mu + \epsilon h_1, \ldots, \mu + \epsilon h_k\}$ are $O(\epsilon^2)$ asymptotically equivalent.*

Proof. We can prove this result using the averaging principle. Let $\bar{\mu}_{\mathrm{A}}$ denote the arithmetic mean of $\{\mu+\epsilon h_1, \ldots, \mu+\epsilon h_k\}$. The geometric mean $\mu_G(\mu + \epsilon h_1, \ldots, \mu + \epsilon h_k) = \left(\prod_{i=1}^k (\mu + \epsilon h_i)\right)^{1/k}$ satisfies the interchangeability and differentiability properties. Therefore, application of Theorem 1 gives

$$\mu_G(\mu+\epsilon h_1, \ldots, \mu+\epsilon h_k) = \mu_G(\bar{\mu}_{\mathrm{A}}, \ldots, \bar{\mu}_{\mathrm{A}})+O(\epsilon^2) = \bar{\mu}_{\mathrm{A}}+O(\epsilon^2).$$

The proof for the harmonic mean $\bar{\mu}_{\mathrm{H}} = k/(\frac{1}{\mu_1} + \cdots + \frac{1}{\mu_k})$ is similar. □

From Lemma 1 and the differentiability of $F_{\mathrm{homog.}}(\mu)$ it follows that

$$\begin{aligned} F_{\mathrm{homog.}}(\bar{\mu}_{\mathrm{A}}) &= F_{\mathrm{homog.}}(\bar{\mu}_{\mathrm{G}}) + O(\epsilon^2) \\ &= F_{\mathrm{homog.}}(\bar{\mu}_{\mathrm{H}}) + O(\epsilon^2). \end{aligned}$$

Corollary 2. *Let $\mu > 0$. Then, the averaging principle (Theorem 1) remains valid if we replace the arithmetic mean with the geometric or harmonic means.*

A natural question is which of the three averages is "optimal", in the sense that it minimizes the constant in the $O(||\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}||^2)$ error term. The answer to this question is model specific. It can be pursued by calculating explicitly the $O(\epsilon^2)$ term, as we will do later on.

Weak Interchangeability. To extend the scope of the averaging principle, we define a weaker interchangeability property.[4]

2A. **Weak interchangeability**: For every $\mu, \tilde{\mu}$ and every $1 \le j_0 \le k$, if $\mu_{j_0} = \tilde{\mu}$ and $\mu_j = \mu$ for all $j \neq j_0$, then $F(\mu_1, \ldots, \mu_k)$ is independent of the value of j_0.

Thus, $F(\mu_1, \cdots, \mu_k)$ is weakly interchangeable if, whenever all but one of the parameters are identical, the outcome F is independent of the identity (coordinate) of the heterogeneous parameter.

Every interchangeable function F is also weakly interchangeable, but not vice versa. Nevertheless, the proof of Theorem 1 implies that:

Corollary 3. *The averaging principle (Theorem 1) remains valid if we replace the assumption of interchangeability with the assumption of weak interchangeability.*

Queuing theory application: An M/M/k queue with heterogeneous service rates

Consider a system with k servers. Server i has a random service time that is distributed according to an exponential distribution with rate μ_i. Customers arrive randomly according to a Poisson distribution with arrival rate λ. An arriving customer is randomly allocated to one of the non-busy servers, if such a server exists. Otherwise, the customer joins a waiting queue, which is unbounded in length. Once a customer is allocated to a server, he gets the service he needs and then leaves the system. This setup is known in the Queuing literature as M/M/k model.[5] Examples for such multi-server queuing systems are call centers, queues in banks, parallel computing, and communications in ISDN protocols.

Let $F(\mu_1, \ldots, \mu_k)$ denote the expected number of customers in the system (i.e., waiting in the queue or receiving service) in steady state. In the case of two heterogeneous servers, $F(\mu_1, \mu_2)$ can be explicitly calculated (see Appendix):

Lemma 2. *Consider an M/M/2 queue with heterogeneous servers. The expected number of customers in the system is given by*

$$F(\mu_1, \mu_2) = \frac{1}{(1-\rho)^2} \frac{1}{\frac{1}{\rho}\frac{2\mu_1\mu_2}{(\mu_1+\mu_2)^2} + \frac{1}{1-\rho}}, \qquad \rho := \frac{\lambda}{\mu_1+\mu_2}. \qquad [4]$$

Finding an explicit solution for $F(\mu_1, \ldots, \mu_k)$ when $k \ge 3$ is computationally challenging, because it involves solving a system of $2^k - 1$ linear equations. In the homogeneous case $\mu_1 = \cdots = \mu_k = \mu$, however, it is well-known that

$$F(\underbrace{\mu, \ldots, \mu}_{\times k}) = \frac{\frac{(\lambda/\mu)^k}{k!}\frac{\frac{\lambda}{k\mu}}{1-\frac{\lambda}{k\mu}}}{\sum_{n=0}^{k-1}\frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^k}{k!}\frac{1}{1-\frac{\lambda}{k\mu}}} \frac{1}{1-\frac{\lambda}{k\mu}} + \frac{\lambda}{\mu}. \quad [5]$$

The function $F(\mu_1, \ldots, \mu_k)$ can be written as a sum of solutions of a system of linear equations with coefficients that depend smoothly on $\mu_1, \ldots, \mu_k$ (see the appendix). Therefore, F is differentiable. Since customers are randomly allocated to the free servers, renaming the servers does not affect the expected number of customers in the system. Hence, F is also interchangeable. Therefore, we can use the averaging principle to obtain an explicit $O(\epsilon^2)$ approximation for $F(\mu_1, \ldots, \mu_k)$:

Theorem 4. *Consider an M/M/k queue with heterogeneous servers whose service rates are $\{\mu_1, \ldots, \mu_k\}$. The expected number of customers in the system is given by*

$$F(\mu_1, \ldots, \mu_k) = F_{\mathrm{homog.}}(\bar{\mu}) + O(\varepsilon^2),$$

where $F_{\mathrm{homog.}}(\bar{\mu}) := F(\underbrace{\bar{\mu}, \ldots, \bar{\mu}}_{\times k})$ is given by [**5**], *$\bar{\mu} := \frac{1}{k}\sum_{i=1}^k \mu_i$, and ϵ is give by* [**2**].

For example, by Theorem 4, the expected number of customers with 2 heterogeneous servers is

$$F(\mu_1, \mu_2) = F(\bar{\mu}, \bar{\mu}) + O\left(\varepsilon^2\right) = \frac{4\lambda\bar{\mu}}{4\,\bar{\mu}^2 - \lambda^2} + O\left(\varepsilon^2\right), \quad [6]$$

where

$$\bar{\mu} = \frac{\mu_1+\mu_2}{2}, \qquad \varepsilon = \frac{\mu_2-\mu_1}{2}.$$

Indeed, substituting $\mu_{1,2} = \bar{\mu} \pm \epsilon$ in the exact expression [**4**] and expanding in ε gives [**6**].

In the case of $k = 8$ heterogeneous servers, even writing the system of $2^8 - 1 = 255$ equations for the 255 unknowns is a formidable task, not to mention solving it explicitly. By the averaging principle, however,

$$F(\mu_1, \ldots, \mu_8) = F_{\mathrm{homog.}}(\bar{\mu}) + O\left(\varepsilon^2\right),$$

where $F_{\mathrm{homog.}}(\bar{\mu}) := F(\underbrace{\bar{\mu}, \ldots, \bar{\mu}}_{\times 8})$ is given by [**5**] with $k = 8$.

We ran stochastic simulations of an M/M/8 queuing system with 8 heterogeneous servers using the ARENA simulation

[4] See the social-networks application below for an example of a weakly-interchangeable outcome which is not interchangeable.

[5] For an introduction to queueing theory, see e.g., [6].

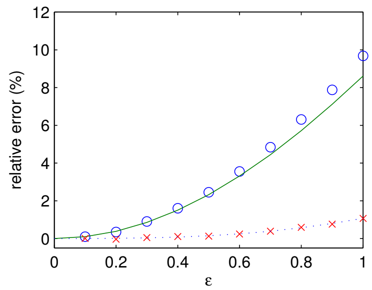

Fig. 1. The relative error of the averaging-principle approximation for the steady-state number of customers in a system with 8 heterogeneous servers, as a function of the heterogeneity parameter ϵ. The solid line is $error = 0.594\epsilon^2$. The crosses denote the relative error of the improved approximation [**15**]. The dotted line is $error = 0.074\epsilon^3$.

software, and used it to calculate the expected number of customers in the system. The simulation parameters were

$$\lambda = \frac{28}{\text{hour}}, \qquad \mu = \frac{5}{\text{hour}}, \quad \mu_i = \mu + \varepsilon h_i, \quad i = 1, \ldots 8,$$

$$(h_1, \ldots, h_8) = (1, 1.5, 2, 3, 3.5, -2.5, -4, -4.5)\frac{1}{\text{hour}},$$

and ε varies between 0 and 1 in increments of 0.05. Because $\sum_{i=1}^{k} h_i = 0$, the average service rate is $\bar{\mu} = \mu = 5$. Therefore, by Theorem 4,

$$F(\mu + \varepsilon h_1, \ldots, \mu + \varepsilon h_8) = F_{\text{homog.}}(5) + O\left(\varepsilon^2\right).$$

In addition, by Equation [**5**], $F_{\text{homog.}}(5) = 6.2314$.

To illustrate the accuracy of this approximation, we plot in Fig. 1 the relative error of the averaging-principle approximation $\frac{F(\mu+\varepsilon h_1,\ldots,\mu+\varepsilon h_8) - F_{\text{homog.}}(5)}{F(\mu+\varepsilon h_1,\ldots,\mu+\varepsilon h_8)}$. As expected, this error scales as ε^2. Note that even when the heterogeneity is not small, the averaging-principle approximation is quite accurate. This is because the coefficient (0.594) of the $O(\epsilon^2)$ term is small.[6] For example, when $\epsilon = 0.5$ the relative error is $\approx 2\%$, and for $\epsilon = 1$ it is below 10%.

Remark. We can also use the averaging principle to obtain $O(\epsilon^2)$ approximations of other quantities of interest that satisfy the interchangeability property, such as the average waiting time in the queue, or the probability that there are exactly m customers in the queue.

Game theory application: Asymmetric Auctions

Consider a sealed-bid first-price auction with k bidders, in which the bidder who places the highest bid wins the object and pays his bid, and all other bidders pay nothing.[7] A common assumption in auction theory is that of independent *private-value auctions*, which says that each bidder knows his own valuation for the object, does not know the valuation of the other bidders, but does know the cumulative distribution functions (CDF) of the valuations of the other bidders. Bidders are also characterized by their attitude towards risk: The literature usually assumes that bidders are risk neutral, since this simplifies the analysis. More often than not, however, bidders are risk averse.

A *strategy* of bidder i is a function $b_i(\cdot)$ that assigns a bid $b_i(v_i)$ to each possible valuation v_i of that bidder. The bid that a bidder places depends on his valuation v_i, and on his beliefs about the distributions of the valuations of the other bidders and about their bidding behavior. An *equilibrium* in this setup is a vector of k strategies $\{b_i(\cdot)\}_{i=1}^{k}$, such that no single bidder can profit by deviating from his bidding strategy, whatever his valuation might be, so long that all other bidders follow their equilibrium bidding strategies.

Most of the auction literature focuses on the *symmetric* (homogeneous) case, in which the beliefs of any bidder about any other bidder (e.g., about his distribution of valuations, his attitude towards risk, etc.) are the same. In this case, one can look for a symmetric equilibrium, in which all bidders adopt the same strategy. In practice, however, bidders are usually asymmetric (heterogeneous), both in their attitude towards risk and in the distribution of their valuations. Each bidder then faces a different competition. As a result, the equilibrium strategies of the bidders are not the same.

The addition of asymmetry usually leads to a huge complication in the analysis. For example, in the case of a first-price auction for a single object with risk-neutral bidders that have private values that are independently distributed in the unit interval $[0, 1]$ according to a common function $F(v)$, the symmetric Nash equilibrium inverse bidding strategy $v(b) = b^{-1}(v)$ satisfies the ordinary differential equation (ODE)[8]

$$v'(b) = \frac{1}{k-1}\frac{F(v(b))}{F'(v(b))}\frac{1}{v(b) - b}, \qquad v(0) = 0.$$

This equation can be solved explicitly, yielding

$$b(v) = v - \frac{\int_0^v F^{k-1}(s)\,ds}{F^{k-1}(v)}. \qquad [\mathbf{7}]$$

Therefore, this case is "completely understood". From the seller's point of view, a key property of an auction is his expected revenue. In the symmetric case, the expression [**7**] can be used to calculate the seller's expected revenue $R_{\text{homog.}}[F]$, yielding

$$R_{\text{homog.}}[F] = 1 + (k-1)\int_0^1 F^k(v)\,dv - k\int_0^1 F^{k-1}(v)\,dv. \quad [\mathbf{8}]$$

In the asymmetric case, where the value of bidder i is independently distributed in $[0, 1]$ according to $F_i(v)$, the inverse equilibrium strategies $\{v_i(\cdot)\}_{i=1}^{k}$ are the solutions of the system of ODE's

$$v_i'(b) = \frac{F_i(v_i(b))}{F_i'(v_i(b))}\left[\left(\frac{1}{k-1}\sum_{j=1}^{k}\frac{1}{(v_j(b) - b)}\right) - \frac{1}{(v_i(b) - b)}\right], \quad [\mathbf{9a}]$$

for $i = 1, \cdots, k$, subject to the initial conditions

$$v_i(b = 0) = 0, \qquad i = 1, \cdots, k, \qquad [\mathbf{9b}]$$

and the "end condition" at some unknown $\bar{b}$

$$v_i(\bar{b}) = 1, \qquad i = 1, \cdots, k. \qquad [\mathbf{9c}]$$

Thus, the addition of asymmetry leads to a huge complication of the mathematical model: instead of a single ODE that can be explicitly integrated, the mathematical model consists of a system of coupled nonlinear ODE's with a non-standard boundary condition. As a result, the system [**9**] cannot be explicitly solved, and it is poorly understood, compared with the symmetric case.

In [2], Fibich and Gavious considered the system [**9**] in the weakly-asymmetric case $F_i = F + \epsilon H_i$, $i = 1, \ldots, k$. After several pages of perturbation-analysis calculations, they obtained

[6] This coefficient will be computed analytically later on from eq. [**14**].

[7] See [7] for an introduction to auction theory.

[8] Since we consider the case where all bidders use the same strategy, we omit the subscript i from b and v.

$O(\epsilon^2)$ asymptotic approximations of the inverse equilibrium strategies $\{v_i(b;\epsilon)\}_{i=1}^k$. Substituting these approximations in the expression for the seller's expected revenue, showed that it is given by

$$R[F_1 = F + \epsilon H_1, \ldots, F_k = F + \epsilon H_k] = R_{\text{homog.}}[F] \quad \textbf{[10]}$$
$$-\epsilon(k-1)\int_0^1 (1-F(v))F^{k-2}(v)\sum_{i=1}^k H_i(v)\,dv + O(\epsilon^2).$$

Subsequently, Lebrun [8] proved that the function on the left-hand-side of **[10]** is differentiable in ϵ, and used that to show that Eq. **[10]** holds. This is, in fact, a special case of the averaging principle. Indeed, interchangeability holds since changing the indices of the bidders does not affect the revenue, and, as mentioned above, differentiability in ϵ was proved in [8]. Therefore, by the averaging principle for functions (see the appendix),

$$R[F_1 = F + \epsilon H_1, \ldots, F_k = F + \epsilon H_k] = R_{\text{homog.}}[\bar{F}] + O(\epsilon^2) \quad \textbf{[11]}$$

where $\bar{F} = F + \frac{\epsilon}{k}\sum_{i=1}^k H_i$. Substituting $\bar{F}$ in **[8]** and expanding in powers of ϵ gives

$$R_{\text{homog.}}[\bar{F}] = R_{\text{homog.}}[F]$$
$$-\epsilon(k-1)\int_0^1 (1-F(v))F^{k-2}(v)\sum_{i=1}^k H_i(v)\,dv + O(\epsilon^2).$$

Hence, relation **[10]** follows.

Numerical calculations ([2, Table 1] and [3, Tables 1-2]) show that the error of the averaging-principle approximation **[11]** is small (typically below 1%), even when the asymmetry level is mild (e.g., $\epsilon = 0.4$). This provides another illustration that the averaging-principle approximation can be useful even when ϵ is not very small.

The averaging principle does not only lead to a simpler derivation of relation **[10]**, but also enables us to derive a more general novel result:

Theorem 5. *Consider an* anonymous *auction*[9] *in which all k bidders have the same attitude towards risk, and all bidders follow the same "rules" when they determine their bidding strategies.*[10] *Let $F_1, \cdots, F_k$ be the cumulative distribution functions of the valuations of the bidders, and let $R[F_1, \cdots, F_k]$ be the expected revenue of the seller. If R is twice differentiable at and near the diagonal, then*

$$R[F_1, \cdots, F_k] = R_{\text{homog.}}[\bar{F}] + O(\epsilon^2),$$

where $R_{\text{homog.}}[\bar{F}] = R[\bar{F}, \cdots, \bar{F}]$, $\bar{F}$ is the average of $F_1, \cdots, F_k$, and ϵ is the level of heterogeneity.

Indeed, the assumptions of the theorem imply that F is interchangeable. Therefore, if F is also differentiable, the theorem follows from the averaging principle.

Social-networks application: Diffusion of new products

Diffusion of new products is a fundamental problem in Marketing, which has been studied in diverse areas such as retail service, industrial technology, agriculture, and educational, pharmaceutical and consumer-durables markets [9]. Typically, the diffusion process begins when the product is first introduced into the market, and progresses through a series of adoption events. An individual can adopt the product due to *external influences* such as mass-media or commercials, and/or due to *internal influences* by other individuals who have already adopted the product (word of mouth). The internal influences depend on the underlying social-network structure, since adopters can only influence people that they "know". The social network is usually modeled by an undirected graph, where each vertex is an individual, and two vertices are connected by an edge if they can influence each other.

The first quantitative analysis of diffusion of new products was the *Bass model* [1], which inspired a huge body of theoretical and empirical research. In this model and in many of the subsequent product-diffusion models:

1. A new product is introduced at time $t = 0$.
2. Once a consumer adopts the product, he remains an adopter at all later times.
3. If consumer j has not adopted before time t, the probability that he adopts the product in the time interval $[t, t+s)$, given that the product was already adopted by $n_j(t)$ people that are connected to j, and that no other consumer adopts the product in the time interval $[t, t+s)$, is

$$\text{Prob}\left(j \text{ adopts in } [t,t+s) \,\middle|\, n_j(t), \begin{array}{c}\text{no other consumer}\\ \text{adopts in } [t,t+s)\end{array}\right)$$
$$= \left(p_j + \frac{n_j(t)}{m_j}\cdot q_j\right)s + O(s^2), \quad \textbf{[12]}$$

as $s \to 0$, where m_j is the total number of individuals connected to consumer j and the parameters p_j and q_j describe the likelihood of individual j to adopt the product due to external and internal influences, respectively.

We say that a social network is *translation invariant*, if any individual sees exactly the same network structure. Therefore, in particular, m_j is independent of j. Examples of translation-invariant social networks are (see Fig. 2):

A) A complete graph, in which any two individuals are connected.
B) A one-dimensional circle, in which each individual is connected to his two nearest neighbors.
C) A one-dimensional circle, in which each individual is connected to his four nearest neighbors.
D) A 2-dimensional torus, in which each individual is connected to his four nearest neighbors.

We say that all individuals are homogeneous when all individuals share the same parameters, i.e., $p_j = p$ and $q_j = q$ for every individual j. Let $N(t)$ denote the number of adopters at time t. The expected aggregate adoption curve $E_{\text{homog.}}[N(t;p,q)]$ in several translation-invariant social networks with homogeneous individuals were analytically calculated in [10, 4]. In these studies, the assumption that all individuals are homogeneous was essential for the analysis.

One of the fundamentals of marketing theory is that consumers are anything but homogeneous. An explicit calculation of the expected aggregate adoption curve

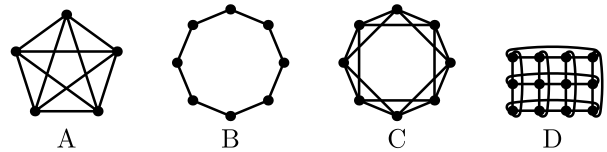

Fig. 2. Examples of translation-invariant networks.

[9] i.e., an auction in which the winner and the amount that each bidder pays depend solely on their bids, and not on the identity of the bidders.

[10] For example, bidders may use *bounded rationality* [11] when determining their bidding strategies. Thus, bidders may restrict themselves to a class of simple strategies, such as low-order polynomial functions of the valuation v. They may even not be aware of the concept of equilibrium. Nevertheless, as long as all bidders have the "same" bounded rationality, the interchangeability requirement holds.

$E[N(t;\{p_j\},\{q_j\})]$ in the heterogeneous case, however, is much harder than in the homogeneous case. As a result, the effect of heterogeneity is not well understood.

The averaging principle allows us to approximate the heterogeneous model with the corresponding homogeneous model. Consider a translation-invariant network. Then, for $t \geq 0$ the function $F(\{p_j\},\{q_j\}) := E[N(t;\{p_j\},\{q_j\})]$ is differentiable and weakly-interchangeable (see Appendix). Therefore, by the averaging principle,

Theorem 6. *The expected aggregate adoption curve in a translation invariant social network with heterogeneous individuals, can be approximated with*

$$E[N(t;\{p_j\},\{q_j\})] = E_{\text{homogeneous}}[N(t;\bar{p},\bar{q})] + O(\epsilon^2),$$

where $\bar{p}$ and $\bar{q}$ are the averages of $\{p_j\}$ and $\{q_j\}$, respectively, and ϵ is the level of heterogeneity of $\{p_j\}$ and $\{q_j\}$.

Theorem 6 is consistent with previous numerical findings:

- In [5], simulations of an agent-based model with a complete graph showed that heterogeneity in p and q had a minor effect on the expected aggregate adoption curve.
- Simulations of agent-based models with 1D and 2D translation-invariant networks [4, Figure 18] showed that when the values of $\{p_j\}$ and $\{q_j\}$ are uniformly distributed within $\pm 20\%$ of the corresponding values $\bar{p}$ and $\bar{q}$ of the homogeneous individuals, the heterogeneous and homogeneous adoption curves are nearly indistinguishable. Even when the heterogeneity level was increased to $\pm 50\%$, the two adoption curves were still very close.

Calculating the $O(\epsilon^2)$ term

The averaging principle is based on a two-term Taylor expansion of F. Therefore, the error of this approximation is given, to leading order, by the quadratic term in this expansion. When F satisfies the differentiability and interchangeability properties[11] and μ is the arithmetic mean, this error is given by (see the appendix):

$$F(\mu_1,\dots,\mu_k) - F(\bar{\mu}_A,\dots,\bar{\mu}_A) \sim \alpha \sum_{i=1}^{k} (\mu_i - \bar{\mu}_A)^2, \quad \textbf{[13a]}$$

where

$$\alpha := \frac{1}{2}\left(\left.\frac{\partial^2 F}{\partial \mu_1 \partial \mu_1}\right|_{\bar{\boldsymbol{\mu}}_A} - \left.\frac{\partial^2 F}{\partial \mu_1 \partial \mu_2}\right|_{\bar{\boldsymbol{\mu}}_A} \right). \quad \textbf{[13b]}$$

Therefore,

1. The magnitude of this error is $\sim |\alpha| \, ||\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}_A||^2$.
2. The sign of this error is the same as the sign of α.

A Taylor expansion in h gives

$$F(\bar{\mu}_A + 2h, \bar{\mu}_A, \underbrace{\bar{\mu}_A,\dots,\bar{\mu}_A}_{\times k-2}) - F(\bar{\mu}_A + h, \bar{\mu}_A + h, \underbrace{\bar{\mu}_A \dots, \bar{\mu}_A}_{\times k-2})$$
$$\sim 2\alpha h^2, \qquad h \ll 1.$$

This shows that in order to determine the sign of α, one can compare the effect of adding h units to two parameters with the corresponding effect of adding $2h$ units to a single parameter.

The value of α can be calculated as follows:

Lemma 3. *Assume that $F(\mu_1,\dots,\mu_k)$ satisfies the differentiability and interchangeability properties. Then,*

$$\alpha = \frac{k}{2(k-1)}\left(\left.\frac{\partial^2 F}{\partial \mu_1 \mu_1}\right|_{\bar{\boldsymbol{\mu}}_A} - \frac{1}{k^2} F''_{\text{homog.}}(\bar{\mu}_A) \right).$$

Therefore, this calculation only requires the explicit calculations of F in the homogeneous case $\mu_1 = \dots = \mu_k$, and in the case that the heterogeneity is limited to a single coordinate (i.e., when $\mu_2 = \dots = \mu_k$). In many cases, this is a considerably easier task than the explicit calculation of F in the fully-heterogeneous case.

To illustrate this, consider again the M/M/k example of Fig. 1 with $k = 8$. While the fully-heterogeneous case requires solving $2^k - 1 = 255$ equations, the single-coordinate heterogeneous case requires solving only $2 \cdot k = 16$ equations. Solving these 16 equations symbolically and using Lemma 3 yields, see the appendix,

$$\alpha(k=8) = \frac{1}{2\lambda\bar{\mu}} \frac{\sum_{i=0}^{12} c_i \left(\frac{\bar{\mu}}{\lambda}\right)^i}{\left(\sum_{i=0}^{7} b_i \left(\frac{\bar{\mu}}{\lambda}\right)^i\right)^2}, \quad \textbf{[14]}$$

where the values of $\{c_i, b_i\}$ are listed in the following table:

i	c_i	b_i
0	1	1
1	45	14
2	999	126
3	14280	840
4	144720	4200
5	1088640	15120
6	6249600	35280
7	27941760	40320
8	97977600	
9	263390400	
10	514382400	
11	653184000	
12	406425600	

In particular, substituting $\bar{\mu} = 5$ and $\lambda = 28$ yields $\alpha \approx 0.00837$. This leads to the improved approximation

$$\begin{aligned} F(\mu_1,\dots,\mu_8) &\approx F_{\text{homog.}}(\bar{\mu}) + \alpha \sum_{i=1}^{8} (\mu_i - \bar{\mu})^2 \\ &\approx F_{\text{homog.}}(5) + 0.594\epsilon^2. \end{aligned} \quad \textbf{[15]}$$

The error of this improved approximation scales as $0.074\epsilon^3$, see Fig. 1, which is the next term in the Taylor expansion. In particular, the relative error of **[15]** is below 1.5% for $0 \leq \epsilon \leq 1$.

Final remarks

The averaging principle is based on a simple observation: the leading-order effects of heterogeneity cancel out when the outcome is interchangeable. Nevertheless, it can lead to a significant simplification of mathematical models in all branches of science. The averaging principle is unrelated to averaging that originates from laws of large numbers in large populations, and it holds, e.g., when there are few servers in a queuing system, or a few bidders in an auction.

The interchangeability and the weak interchangeability properties are usually easy to check. The differentiability of F is easy to check in some cases, but can be quite a challenge in others. We note, however, that more often than not, functions that arise in mathematical models are differentiable, unless

[11] Here we cannot assume that F is only weakly interchangeable, since we require that $\frac{\partial^2 F}{\partial \mu_i \partial \mu_j} = \frac{\partial^2 F}{\partial \mu_1 \partial \mu_2}$ for all i, j.

there is a "very good reason" why they are not. While this is a very informal statement, we make it in order to point out that the "generic" case is that the outcome F is differentiable, rather then the other way around.

An important issue is the "level of heterogeneity" that is covered by the averaging principle. Strictly speaking, the level of heterogeneity should be "sufficiently small". In practice, however, in many cases the averaging principle provides good approximations even when $\epsilon = 0.5$. In other words, the coefficient of the $O(\epsilon^2)$ term is O(1). While this is also an informal statement, we make it in order to point out that one should not be "surprised" that the averaging principle holds even when ϵ is not very small.

ACKNOWLEDGMENTS. We thank Uri Yechiali for suggesting the problem of M/M/k queues with heterogeneous servers and the solution of the M/M/2 problem. The research of Solan was partially supported by the ISF grant #212/09 and by the Google Inter-university center for Electronic Markets and Auctions.

1. Bass, F.M., A New Product Growth Model for Consumer Durables. Management Science 15, pp. 215-227(1969).
2. Fibich, G. and Gavious, A., Asymmetric first-price auctions - a perturbation approach, Mathematics of Operational Research, 28, 836852 (2003).
3. Fibich, G. Gavious, A., and Sela, A., Revenue Equivalence in Asymmetric Auctions, Journal of Economic Theory, 115, 309-321 (2004).
4. Fibich, G. and Gibori R., Aggregate diffusion dynamics in agent-based models with a spatial structure, Operations Research, 58,1450-1468 (2010).
5. Goldenberg, J., Libai, B. and Muller, E., Using Complex Systems Analysis to Advance Marketing Theory Development. Academy of Marketing Science Review. [Online] 01, special issue on Emergent and Co-evolutionary Processes in Marketing (2001).
6. Gross, D. and Harris, C. M., A Fundamentals of Queueing Theory, 2nd ed., John Wiley & Sons, New York, (1985).
7. Krishna, V. Auction Theory, Academic Press, San Diego (2002).
8. Lebrun, B., Auctions with Almost Homogeneous Bidders, Journal of Economic Theory, 144, 1341-1351 (2009).
9. Mahajan, V., Muller, E. and Bass, F. M., New-Product Diffusion Models. In J. Eliashberg and G. L. Lilien, editors, Handbooks in Operations Research and Management Science, Volume 5: Marketing, pp. 349-408. North-Holland, Amsterdam, The Netherlands (1993).
10. Niu, S.C., A Stochastic Formulation of the Bass Model of New-Product Diffusion. Mathematical Problems in Engineering 8: 249-263 (2002).
11. Selten, R., Features of experimentally observed bounded rationality, European Economic Review, 42, 413-436 (1988).

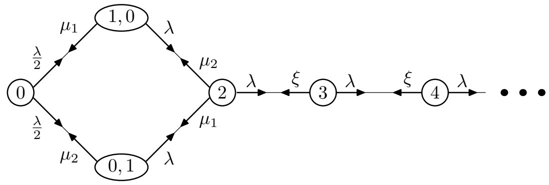

Fig. 3. Transition diagram of a queue with two heterogeneous servers. State "0" corresponds to the situation in which no server is busy. States $(1,0)$ corresponds to the situation in which server 1 is busy and server 2 is not busy. States $(0,1)$ corresponds to the situation in which server 1 is not busy and server 2 is busy. State "k" for $k \geq 2$ corresponds to the situation in which both servers are busy and $k-2$ customers wait in the queue.

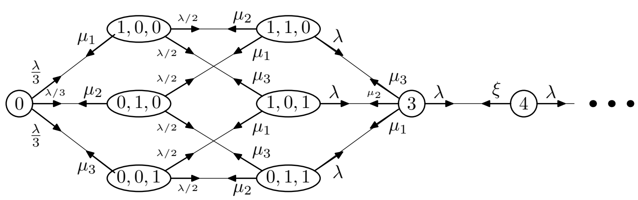

Fig. 4. Same as Fig. 3 with three heterogeneous servers. For example, state $(0,1,1)$ corresponds to the situation in which server 1 is not busy and servers 2 and 3 are busy.

Appendix

Proof of Theorem 1

Because of the differentiability of F, there exists a positive constant $C_{\bar{\mu}_A}$, such that for all $||\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}_A|| < C_{\bar{\mu}}$, we can expand $F(\boldsymbol{\mu})$ as

$$F(\boldsymbol{\mu}) = F(\bar{\boldsymbol{\mu}}_A) + \sum_{j=1}^{k} (\mu_j - \bar{\mu}_A) \frac{\partial F}{\partial \mu_j}\bigg|_{\bar{\boldsymbol{\mu}}_A} + O(||\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}_A||^2).$$

Because F is interchangeable,

$$\frac{\partial F}{\partial \mu_i}\bigg|_{\bar{\boldsymbol{\mu}}} = \frac{\partial F}{\partial \mu_1}\bigg|_{\bar{\boldsymbol{\mu}}_A}, \qquad j = 1, \ldots, k.$$

Therefore,

$$F(\boldsymbol{\mu}) = F(\bar{\boldsymbol{\mu}}_A) + \frac{\partial F}{\partial \mu_1}\bigg|_{\bar{\boldsymbol{\mu}}_A} \sum_{j=1}^{k} (\mu_j - \bar{\mu}_A) + O(||\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}_A||^2).$$

Since $\bar{\mu}_A$ is the arithmetic average, $\sum_{j=1}^{n} (\mu_j - \bar{\mu}_A) = 0$. Hence, the result follows.

Proof of Lemma 2

We calculate $F(\mu_1, \mu_2)$ explicitly using the steady-state transition diagram that is shown in Fig. 3. We denote by p_i the steady-state probability for the system to be with i customers, and by $p_1^{(1,0)}$ and $p_1^{(0,1)}$ the steady-state probability for the system to be with 1 customer in server 1 and 2, respectively. In particular, $p_1 = p_1^{(1,0)} + p_1^{(0,1)}$. Since in steady state the amount of inflow is equal to the amount of outflow, the following equalities hold:

$$\lambda p_0 = \mu_1 p_1^{(1,0)} + \mu_2 p_1^{(0,1)}, \qquad \textbf{[16a]}$$

$$\frac{\lambda}{2} p_0 + \mu_2 p_2 = (\lambda + \mu_1) p_1^{(1,0)}, \qquad \textbf{[16b]}$$

$$\frac{\lambda}{2} p_0 + \mu_1 p_2 = (\lambda + \mu_2) p_1^{(0,1)}, \qquad \textbf{[16c]}$$

$$\lambda p_1^{(1,0)} + \lambda p_1^{(0,1)} + (\mu_1 + \mu_2) p_3 = (\lambda + \mu_1 + \mu_2) p_2, \qquad \textbf{[16d]}$$

$$\begin{aligned} &\lambda p_n + (\mu_1 + \mu_2) p_{n+2} \\ &\quad = (\lambda + \mu_1 + \mu_2) p_{n+1}, \quad n = 2, 3, \ldots \end{aligned} \qquad \textbf{[16e]}$$

We can view **[16a]**–**[16c]** as a linear system for the three unknowns $\{p_0, p_1^{(1,0)}, p_1^{(0,1)}\}$. Solving this system for p_0 yields

$$p_0 = \frac{2\mu_1 \mu_2}{\lambda^2} p_2.$$

In addition, the solution of **[16d]**–**[16e]** is $p_n = \left(\frac{\lambda}{\mu_1 + \mu_2}\right)^{n-2} p_2 = \rho^{n-2} p_2$ for $n \geq 1$. Substituting the above in

$$1 = \sum_{n=0}^{\infty} p_n = p_0 + \sum_{n=1}^{\infty} \rho^{n-2} p_2 = \left(\frac{2\mu_1\mu_2}{\lambda^2} + \frac{1}{\rho}\frac{1}{1-\rho}\right) p_2,$$

gives $p_2 = \left(\frac{2\mu_1\mu_2}{\lambda^2} + \frac{1}{\rho}\frac{1}{1-\rho}\right)^{-1}$. Therefore,

$$\begin{aligned} F(\mu_1, \mu_2) &= \sum_{n=0}^{\infty} n p_n = \sum_{n=0}^{\infty} n \rho^{n-2} p_2 = \frac{p_2}{\rho} \sum_{n=0}^{\infty} n \rho^{n-1} \\ &= \frac{p_2}{\rho} \left(\sum_{n=0}^{\infty} \rho^n\right)' = \frac{p_2}{\rho}\left(\frac{1}{1-\rho}\right)' = \frac{p_2}{\rho}\frac{1}{(1-\rho)^2}, \end{aligned}$$

and the result follows.

M/M/3 queue

Consider the case of three heterogeneous servers with average service times μ_1, μ_2 and μ_3. Denote by p_0, $p_1^{(1,0,0)}$, $p_1^{(0,1,0)}$, $p_1^{(0,0,1)}$, $p_2^{(1,1,0)}$, $p_2^{(1,0,1)}$, $p_2^{(0,1,1)}$, p_3, p_4, …, the steady-state probabilities. Thus, for example, $p_2^{(1,0,1)}$ is the steady-state probability that servers 1 and 3 are busy, server 2 is free, and there are no waiting customers in the queue (we denote by $p_n, n \geq 2$ the probability having n customers in the system). The transition diagram for $k = 3$ servers is given in Fig. 4. The steady-state equations are

$$\begin{aligned} &\lambda p_0 = \mu_1 p_1^{(1,0,0)} + \mu_2 p_1^{(0,1,0)} + \mu_3 p_1^{(0,0,1)}, \\ &\frac{\lambda}{3} p_0 + \mu_2 p_2^{(1,1,0)} + \mu_3 p_2^{(1,0,1)} = (\mu_1 + \lambda) p_1^{(1,0,0)}, \\ &\frac{\lambda}{3} p_0 + \mu_1 p_2^{(1,1,0)} + \mu_3 p_2^{(0,1,1)} = (\mu_2 + \lambda) p_1^{(0,1,0)}, \\ &\frac{\lambda}{3} p_0 + \mu_1 p_2^{(1,0,1)} + \mu_2 p_2^{(0,1,1)} = (\mu_3 + \lambda) p_1^{(0,0,1)}, \\ &\frac{\lambda}{2} p_1^{(1,0,0)} + \frac{\lambda}{2} p_1^{(0,1,0)} + \mu_3 p_3 = (\lambda + \mu_1 + \mu_2) p_2^{(1,1,0)}, \\ &\frac{\lambda}{2} p_1^{(1,0,0)} + \frac{\lambda}{2} p_1^{(0,0,1)} + \mu_2 p_3 = (\lambda + \mu_1 + \mu_3) p_2^{(1,0,1)}, \\ &\frac{\lambda}{2} p_1^{(0,1,0)} + \frac{\lambda}{2} p_1^{(0,0,1)} + \mu_1 p_3 = (\lambda + \mu_2 + \mu_3) p_2^{(0,1,1)}, \\ &\quad = (\lambda + \mu_1 + \mu_2 + \mu_3) p_3, \\ &\lambda p_n + (\mu_1 + \mu_2 + \mu_3)\, p_{n+2} \\ &\quad = (\lambda + \mu_1 + \mu_2 + \mu_3) p_{n+1}, \qquad n \geq 3, \\ &\sum_{n=0}^{\infty} p_n = 1. \end{aligned}$$

The solution of the last two equations is $p_n = \left(\frac{\lambda}{\mu_1+\mu_2+\mu_3}\right)^{n-3} p_3$ for $n \geq 2$. The values of $\{p_0, p_1, p_2\}$ as a function of p_3 can be evaluated explicitly with MAPLE, by solving the first $2^3 - 1 = 7$ linear equations for $\{p_0,\ p_1^{(1,0,0)},\ p_1^{(0,1,0)},\ p_1^{(0,0,1)},\ p_2^{(1,1,0)},\ p_2^{(1,0,1)},\ p_2^{(0,1,1)}\}$. The resulting expression for $F(\mu_1, \mu_2, \mu_3)$, however, is extremely cumbersome and not informative.

Proof of Theorem 4

Since customers are randomly assigned to the available servers, $F(\mu_1, \dots, \mu_k)$ is interchangeable. To see that F is differentiable in $(\mu_1, \dots, \mu_k)$, we note that $F = \sum_{n=0}^{\infty} n p_n$ where p_n is the steady-state probability that there are n customers in the system. In addition, $\{p_n\}_{n=1}^{k}$ are the solutions of a linear system with coefficients that depend smoothly on $(\mu_1, \dots, \mu_k)$, and $p_n = (\frac{\lambda}{\mu_1+\cdots+\mu_k})^{n-k} p_k$ for $n \geq k-1$. This was shown explicitly for the cases $k = 2$ and $k = 3$; the proof for $k > 3$ is similar.

Averaging principle for functions (proof of eq. [11])

Let $(F_j(\mathbf{x}))_{j=1}^{k}$ be functions in the same function space $\mathcal{F}$, and let $\epsilon \in \mathbb{R}$. Let $R : (F_1, \dots, F_k) \mapsto R[F_1, \dots, F_k] \in \mathbb{R}$ be a functional. We say that the functional R is interchangeable, if $R(\dots, F_i, \dots, F_j, \dots) = R(\dots, F_j, \dots, F_i, \dots)$ for all $i \neq j$. We say that the functional R is differentiable if the scalar function $\tilde{R}(\epsilon) := R[F_1 = F + \epsilon H_1, \dots, F_k = F + \epsilon H_k]$ is twice differentiable at and near $\epsilon = 0$, for every $F \in \mathcal{F}$ and every $(H_j(\mathbf{x}))_{j=1}^{k} \in \mathcal{F}^k$.

Given functions $(F_j(\mathbf{x}))_{j=1}^{k}$ in $\mathcal{F}$, denote $\bar{F} = \frac{1}{k}\sum_{j=1}^{k} F_j$ and $H_j = F_j - \bar{F}$. By Taylor expansion,

$$\tilde{R}(\epsilon) = \tilde{R}(0) + \epsilon \sum_{j=1}^{k} \frac{\delta R}{\delta F_j} H_j + O(\epsilon^2),$$

where $\frac{\delta R}{\delta F_j}$ is the variational derivative. Because R is interchangeable,

$$\tilde{R}(\epsilon) = \tilde{R}(0) + \epsilon \frac{\delta R}{\delta F_1} \sum_{j=1}^{k} H_j + O(\epsilon^2).$$

In particular, if $F = \bar{F}$, then $\sum_{j=1}^{k} H_j = 0$. Hence, $\tilde{R}(\epsilon) = \tilde{R}(0) + O(\epsilon^2)$, which is **[11]**.

Proof of Theorem 6

We first prove that F is differentiable. Denote $\delta_{i,i'} = 1$ if individuals i and i' influence each other, and $\delta_{i,i'} = 0$ otherwise. For every k, every set of k consumers $\{i_1, i_2, \dots, i_k\}$, and every increasing sequence of times $0 \leq t_1 \leq \dots \leq t_k$, denote by $P(i_1, t_1, i_2, t_2, \dots, i_k, t_k)$ the probability that consumer i_1 adopts the product before time t_1, consumer i_2 adopts the product between times t_1 and t_2, etc., and all consumers who are not in $\{i_1, \dots, i_k\}$ do not adopt the process by time t_k. Then,

$$P(i_1, t_1) = \big(1 - \exp(-p_{i_1} t_1)\big) \prod_{j \neq i_1} \exp(-p_j t_1).$$

Similarly,

$$\begin{aligned} &P(i_1, t_1, i_2, t_2, \dots, i_k, t_k) = \\ &\quad P(i_1, t_1, i_2, t_2, \dots, i_{k-1}, t_{k-1}) \\ &\times \left(1 - \exp\left(-\Big(p_{i_k} + \sum_{m=1}^{k-1} \delta_{i_k, i_m} q_{i_m}\Big)(t_k - t_{k-1})\right)\right) \\ &\times \prod_{j \notin \{i_1, \dots, i_k\}} \exp\left(-\left(p_j + \sum_{m=1}^{k-1} \delta_{j, i_m} q_{i_m}\right)(t_k - t_{k-1})\right). \end{aligned}$$

Hence, the function $P(i_1, t_1, i_2, t_2, \dots, i_k, t_k)$ is differentiable in $\{p_i, q_i\}$. Finally,

$$E[N(t; \{p_j\}, \{q_j\})] = \frac{1}{M} \sum_{\pi} \sum_{k=1}^{M} \frac{k}{(M-k!}$$

$$\times \int_{t_1=0}^{t} \int_{t_2=t_1}^{t} \cdots \int_{t_{k-1}=t_{k-2}}^{t} P(i_1, t_1, \dots, i_{k-1}, t_{k-1}, i_k, t)\, dt_{k-1} \dots dt_1,$$

where π ranges over all permutations on the set of M individuals. Therefore, the differentiability of $E[N(t; \{p_j\}, \{q_j\})]$ follows.

Because the network is translation invariant, F is weakly-interchangeable in $\{p_j\}$ and in $\{q_j\}$. By this we mean that

- If $p_m = \tilde{p}$, $p_j = p$ for all $j \neq m$, and $q_j = q$ for all j, then F is independent of the value of m.
- If $q_n = \tilde{q}$, $q_j = q$ for all $j \neq n$, and $p_j = p$ for all j, then F is independent of the value of n.

Therefore, the result follows from a slight modification of the proof of Theorem 1.

Proof of equation [13]

Since F is interchangeable, the quadratic term in the Taylor expansion of $F(\mu_1, \dots. \mu_k)$ around the arithmetic mean is equal to

$$\sum_{i,j=1}^{k} (\mu_i - \bar{\mu}_A)(\mu_j - \bar{\mu}_A) \frac{\partial^2 F}{\partial \mu_i \partial \mu_j}\bigg|_{\bar{\boldsymbol{\mu}}_A} =$$

$$\frac{\partial^2 F}{\partial \mu_1 \partial \mu_2}\bigg|_{\bar{\boldsymbol{\mu}}} \sum_{i,j=1, i \neq j}^{k} (\mu_i - \bar{\mu}_A)(\mu_j - \bar{\mu}_A) + \frac{\partial^2 F}{\partial \mu_1 \partial \mu_1}\bigg|_{\bar{\boldsymbol{\mu}}} \sum_{i=1}^{k} (\mu_i - \bar{\mu}_A)^2.$$

Since $\bar{\mu}_A$ is the arithmetic mean,

$$\sum_{i,j=1}^{k} (\mu_i - \bar{\mu}_A)(\mu_j - \bar{\mu}_A) = \sum_{i=1}^{k} (\mu_i - \bar{\mu}_A) \sum_{j=1}^{k} (\mu_j - \bar{\mu}_A) = 0.$$

Therefore, the result follows.

Proof of Lemma 3

Consider the case where $\mu_i = \bar{\mu} + h$ for $i = 1, \dots, k$. By equation **[13]**,

$$\begin{aligned} &\frac{1}{2} \sum_{i,j=1}^{k} (\mu_i - \bar{\mu})(\mu_j - \bar{\mu}) \frac{\partial^2 F}{\partial \mu_i \partial \mu_j}\bigg|_{\bar{\boldsymbol{\mu}}} \\ &\quad = \frac{1}{2} \frac{\partial^2 F}{\partial \mu_1 \partial \mu_2}\bigg|_{\bar{\boldsymbol{\mu}}} k(k-1)h^2 + \frac{1}{2} \frac{\partial^2 F}{\partial \mu_1 \partial \mu_1}\bigg|_{\bar{\boldsymbol{\mu}}} k h^2. \end{aligned}$$

On the other hand, since

$$F(\bar{\mu} + h, \dots, \bar{\mu} + h) = F_{\text{homog.}}(\bar{\mu} + h),$$

we have

$$\frac{1}{2} \sum_{i,j=1}^{k} (\mu_i - \bar{\mu})(\mu_j - \bar{\mu}) \frac{\partial^2 F}{\partial \mu_i \partial \mu_j}\bigg|_{\bar{\boldsymbol{\mu}}} = \frac{h^2}{2} F''_{\text{homog.}}(\bar{\mu}).$$

Therefore,

$$\frac{1}{2}\frac{\partial^2 F}{\partial\mu_1\partial\mu_2}\bigg|_{\bar{\boldsymbol{\mu}}} k(k-1)h^2 + \frac{1}{2}\frac{\partial^2 F}{\partial\mu_1\partial\mu_1}\bigg|_{\bar{\boldsymbol{\mu}}} kh^2 = \frac{h^2}{2}F''_{\text{homog.}}(\bar{\mu}).$$

Hence,

$$\frac{\partial^2 F}{\partial\mu_1\partial\mu_2}\bigg|_{\bar{\boldsymbol{\mu}}} = \frac{1}{k-1}\left(\frac{1}{k}F''_{\text{homog.}}(\bar{\mu}) - \frac{\partial^2 F}{\partial\mu_1\partial\mu_1}\bigg|_{\bar{\boldsymbol{\mu}}}\right).$$

Calculation of α

We illustrate the computation of the coefficient α for a queue with 8 servers. Consider then the case of a single server with service time μ_1, and seven servers with service time μ. Denote by $p_{0,n}$ and $p_{1,n}$, $n = 1,\dots,6$, the steady-state probabilities that n out of the homogeneous servers are busy and that the single heterogeneous servers is free or busy, respectively. The equations for the $2\cdot 8 - 1 = 15$ variables $\{p_0, p_{0,1}, p_{1,0},\dots,p_{1,6}, p_{0,7}\}$ are

$$\begin{aligned}
&\lambda p_{0,0} = \mu p_{0,1} + \mu_1 p_{1,0}\ ,\\
&-p_{0,0}\frac{\lambda}{8} + p_{1,0}(\lambda+\mu_1) - p_{1,1}\mu = 0,\\
&p_{0,n}\,(\lambda+n\mu) = p_{0,n-1}\frac{8-n}{9-n}\lambda + p_{1,n}\mu_1 + p_{0,n+1}(n+1)\mu,\\
&\quad n = 1,\dots,6,\\
&p_{1,n}\,(\mu_1+\lambda+n\mu) = p_{1,n-1}\lambda + p_{1,n+1}(n+1)\mu + p_{0,n}\frac{\lambda}{8-n},\\
&\quad n = 1,\dots,5,\\
&p_{0,7}(\lambda+7)\mu) = p_{0,6}\frac{\lambda}{2} + p_7\mu_1\rho,
\end{aligned}$$

where $\rho = \frac{\lambda}{7\mu+\mu_1}$, $p_n = \rho^{n-7}p_7$ for $n \geq 8$, and $\sum_{n=0}^{\infty} p_n = 1$. These equations can be solved with Maple [12], and the solution can be used to calculate $F(\mu_1, \underbrace{\mu,\dots,\mu}_{\times 7})$ explicitly. Differentiating this expression twice with respect to μ_1, differentiating $F_{\text{homog.}}$, see eq. **[5]**, twice with respect to μ, and using Lemma 3, yields equation **[14]**. Substituting $\bar{\mu} = 5$ and $\lambda = 28$ gives $\alpha \approx 0.00837$. In addition, $\sum_{i=1}^{8}(\mu_i - \bar{\mu})^2 = \epsilon^2\sum_{i=1}^{8} h_i^2 = 71\epsilon^2$. Therefore, $\alpha\sum_{i=1}^{8}(\mu_i-\bar{\mu})^2 \approx 0.594\epsilon^2$.

[12] The Maple code is available at `www.bgu.ac.il/~ariehg/averagingprinciple.html`.