

SCIENTIFIC REPORTS



OPEN

First passage events in biological systems with non-exponential inter-event times

Mario Castro^{1,2}, Martín López-García², Grant Lythe² & Carmen Molina-París²

It is often possible to model the dynamics of biological systems as a series of discrete transitions between a finite set of observable states (or *compartments*). When the residence times in each state, or inter-event times more generally, are exponentially distributed, then one can write a set of ordinary differential equations, which accurately describe the evolution of mean quantities. Non-exponential inter-event times can also be experimentally observed, but are more difficult to analyse mathematically. In this paper, we focus on the computation of first passage events and their probabilities in biological systems with non-exponential inter-event times. We show, with three case studies from Molecular Immunology, Virology and Epidemiology, that significant errors are introduced when drawing conclusions based on the assumption that inter-event times are exponentially distributed. Our approach allows these errors to be avoided with the use of phase-type distributions that approximate arbitrarily distributed inter-event times.

Biology provides many instances of processes where dynamics far from equilibrium determines the fate of the system. Notable examples include the ability of a virus to infect a cell and produce new virions before the immune system detects it, or the initiation of a signalling cascade after ligand-receptor binding and phosphorylation events¹. Many mathematical models of these biological systems have been developed at the population level and make use of ordinary differential equations. However, new experimental techniques that allow study of single individual agents (cells, receptors, genes, ...) demand stochastic descriptions² of these biological processes. Stochastic methods are growing in popularity, partly driven by new questions posed by more sophisticated experiments, but also by the increasing capacity of modern computers. The common Markovian assumption in stochastic models is that events in the future, given information only about the present state of the system, are independent of those that occurred in the past (the so-called memoryless property). A particular property of continuous time Markov processes with discrete state spaces (*i.e.*, continuous time Markov chains (CTMCs)) is that the time for each event to occur is exponentially distributed. Thus, the next event to occur in the CTMC is the result of a competition between exponentially-distributed *absolute waiting times*, leading to the Gillespie algorithm³ for their efficient numerical simulation.

A stochastic process in which inter-event times are not exponentially distributed is not strictly Markovian because, as well as depending on the current state of the system, the choice and timing of the next event depend on how long the process has been in its current state. It is not difficult to find experimental situations where the Markovian assumption is too strong. For instance, the probability density of the time from infection of a cell until the same cell produces a new virion is not exponentially distributed⁴; multiple intermediate steps (virus entry, migration to the nucleus, hijacking of the cellular machinery, viral assembly, ...) take place inside the cell between these two events. Another biological process where it has been suggested that the Markovian assumption might be too strong is cell proliferation⁵, where Erlang distributions for modelling the inter-event times within the cell cycle might be more appropriate than the exponential distribution⁵. It is also important to keep in mind that, when more than one transition is possible out of a given state, experimentally observed distributions of times for a particular transition to occur are always conditioned (also referred to as censored⁶) on the chosen transition actually taking place.

In this paper, we relax the Markovian assumption by requiring that the process is Markovian only at arrival times: once the process jumps into any state, the future of the system does not depend on its past, but only on how

¹Grupo Interdisciplinar de Sistemas Complejos (GISC) and DNL, Universidad Pontificia Comillas, Madrid, E-28015, Spain. ²Department of Applied Mathematics, School of Mathematics, University of Leeds, Leeds, LS2 9JT, UK. Correspondence and requests for materials should be addressed to M.C. (email: marioc@comillas.edu)

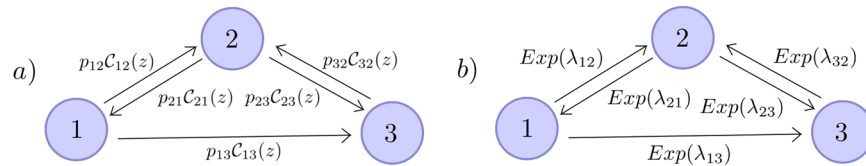


Figure 1. (a) Flowgraph representation of a continuous time semi-Markov process, \mathcal{X} , defined on $\mathcal{S} = \{1, 2, 3\}$. (b) Markovian version of the process \mathcal{X} , where all events $i \rightarrow j$, with $i, j \in \mathcal{S}$, occur with exponentially distributed times $T_{i \rightarrow j} \sim \text{Exp}(\lambda_{ij})$.

long the process has been in that state, leading to a *semi-Markov* process⁷. Although Markov processes have been widely studied in the literature, and their applicability has been shown in a wide range of areas such as Biology, Epidemiology, Immunology, Population Dynamics or Queuing Theory, semi-Markov processes have mostly been treated as an advanced and theoretical topic with a focus on definitions and limit theorems⁸, but with limited applications^{9,10}. A practical way of analysing semi-Markov processes, and linking them to real data, is the “flowgraph” methodology⁸. Flowgraphs are elegant graphical representations of semi-Markov processes: the system is depicted as a directed network formed by the possible states of the process, where arrows represent possible events (or transitions) causing the process to move (or *jump*) from its present state to another (see Fig. 1a). Each event has an associated probability and conditioned waiting time. It is convenient to describe these conditioned waiting times by their moment generating functions (or Laplace-Stieltjes transforms), as in Fig. 1a. One of the advantages of flowgraphs is that they bypass the required mathematical characterisation of semi-Markov processes, since they are designed to handle real data⁸.

The aim when analysing flowgraph models is usually to study probabilities of potential outcomes (*e.g.*, the probability that the process in Fig. 1a reaches state 3 before reaching state 2, if the process starts at 1), or first passage times (*e.g.*, the time for the process to reach state 3, if the process starts at 1). In this paper, we focus on computing a first passage time (the time for the process to reach a given state) when some inter-event times are not exponentially distributed. First passage times have been widely analysed in the literature when considering stochastic systems with continuous states and with specific applications in Biology^{11,12}. In this manuscript, we focus on computing first passage times in systems with discrete states. For discrete states, first passage times have been successfully analysed in the Markovian case¹³ [Chapter 1], and some work has already been carried out within the theory of semi-Markov processes but under very restrictive assumptions¹⁴ and with a strong theoretical focus. When considering the flowgraph methodology as a way of analysing semi-Markov processes, first passage times can be analysed but it is usual to assume that the probabilities of each event occurring in the system, as well as the conditioned waiting time distributions, are known⁸.

The purpose of this paper is threefold:

1. to obtain the relationship between absolute and conditioned waiting times when all or some inter-event times are not exponentially distributed,
2. to introduce a methodology based on phase-type distributions to approximate non-Markovian events¹³, while maintaining analytical tractability, and
3. to show that considering (*e.g.*, biological) processes with the inappropriate assumption that all inter-event times are exponentially distributed can lead to incorrect predictions (*e.g.*, when analysing first passage times).

We introduce our results using three case studies from different biological areas. We summarise the main mathematical results in the Materials and Methods section, and devote most of the Results section, to illustrating their applicability in three examples taken from Immunobiology (at the receptor level), Virology (at the cell level) and Epidemiology (at the host level). Further details of the derivation and additional results may be found in the Supplementary Information.

Materials and Methods

The flowgraph methodology. We consider a continuous time stochastic process $\mathcal{X} = \{X(t): t \geq 0\}$ defined over a finite and discrete space of states, \mathcal{S} . Given any state $i \in \mathcal{S}$, we define $A(i) = \{i_1, i_2, \dots, i_n\}$ to be the set of states directly accessible (*i.e.*, in one *jump*) from i , where n_i is the cardinality of $A(i)$. For example, for the process \mathcal{X} in Fig. 1, we have $A(1) = \{2, 3\}$, $A(2) = \{1, 3\}$ and $A(3) = \{2\}$.

If the process \mathcal{X} starts at $t = 0$ in state $i \in \mathcal{S}$, $X(0) = i$, it jumps from i to any other state $j \in A(i)$ after a random time. This event (or transition) $i \rightarrow j$ actually occurs, if and only if, the *absolute waiting time* $T_{i \rightarrow j}$ for this event is shorter than any other *competing* absolute waiting time $T_{i \rightarrow j'}, j' \in A(i) \setminus \{j\}$. This means that the transition probability of the event $i \rightarrow j$ can be written as

$$p_{ij} = \mathbb{P}(i \rightarrow j) = \mathbb{P}(T_{i \rightarrow j} < T_{i \rightarrow j'}, \forall j' \in A(i) \setminus \{j\}), j \in A(i).$$

The time for this transition to take place, conditioned on actually occurring, $T_{i \rightarrow j} | i \rightarrow j$ (*i.e.*, the *conditioned waiting time*), can be described in terms of its Laplace-Stieltjes transform

$$C_{ij}(z) = E[e^{-zT_{i \rightarrow j}} | i \rightarrow j].$$

We assume that the process is Markovian at arrival instants. Namely, after the arrival at a given state of the process $i \in \mathcal{S}$, future transitions do not depend on the past history of the process. However, we note that once the process has been in a given state for some time, its future evolution might, in fact, depend on how long the process has been in that state, so that the process is only Markovian at arrival to any given state, and is not a Markov process, but a semi-Markov process⁷. A flowgraph representation⁸ [Chapter 2] of \mathcal{X} is a set of nodes (states) connected by arrows that represent the possible transitions between states. The arrows are labelled with transition probabilities, p_{ij} , and conditioned Laplace-Stieltjes transforms, $C_{ij}(z)$, as defined above (see the example in Fig. 1a). We note that the Laplace-Stieltjes transforms can be replaced by the moment generating functions, $M_{ij}(s) = E[e^{sT_{i \rightarrow j}} | i \rightarrow j] = C_{ij}(-s)$ for $i, j \in \mathcal{S}$. Below, we will use the notation of Fig. 1b, where each arrow is accompanied by its absolute waiting time distribution. The reader can translate between both notations using Fig. 1.

Let us imagine that the process \mathcal{X} is defined on the space of states $\mathcal{S} = \{1, 2, \dots, N\} \cup \{\Omega\}$, with initial state $X(0) = i \in \mathcal{S}$, and that the interest is in studying the first passage time to state Ω , that is, $T_i(\Omega) = \inf\{t \geq 0: X(t) = \Omega\}$. This time can be analysed in terms of its Laplace-Stieltjes transform, $\mathcal{L}_i(z) = E[e^{-zT_i(\Omega)}]$, which can be computed by carrying out a first step argument (we can assume that, under the semi-Markov property, only the current state affects the next transition), as follows:

$$\begin{aligned} \mathcal{L}_i(z) &= \sum_{j \in A(i)} E[e^{-zT_i(\Omega)} | i \rightarrow j] \mathbb{P}(i \rightarrow j) \\ &= \sum_{j \in A(i)} E[e^{-z(T_{i \rightarrow j} + T_j(\Omega))} | i \rightarrow j] \mathbb{P}(i \rightarrow j) \\ &= \sum_{j \in A(i)} E[e^{-zT_{i \rightarrow j}} | i \rightarrow j] E[e^{-zT_j(\Omega)} | i \rightarrow j] \mathbb{P}(i \rightarrow j) \\ &= \sum_{j \in A(i)} p_{ij} C_{ij}(z) \mathcal{L}_j(z), \end{aligned}$$

which leads to the following system of equations in matrix form⁹

$$\begin{pmatrix} 1 & -p_{12}C_{12}(z) & \dots & -p_{1N}C_{1N}(z) & -p_{1\Omega}C_{1\Omega}(z) \\ -p_{21}C_{21}(z) & 1 & \dots & -p_{2N}C_{2N}(z) & -p_{2\Omega}C_{2\Omega}(z) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -p_{N1}C_{N1}(z) & -p_{N2}C_{N2}(z) & \dots & 1 & -p_{N\Omega}C_{N\Omega}(z) \\ 0 & 0 & \dots & \dots & 1 \end{pmatrix} \begin{pmatrix} \mathcal{L}_1(z) \\ \mathcal{L}_2(z) \\ \vdots \\ \mathcal{L}_N(z) \\ \mathcal{L}_\Omega(z) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \tag{1}$$

with $\mathcal{L}_\Omega(z) \equiv 1$ (as the time to reach Ω starting at Ω is 0). Once the Laplace-Stieltjes transform $\mathcal{L}_i(z)$ has been determined for the particular initial state of interest $i \in \mathcal{S}$, the moments of $T_i(\Omega)$ can be calculated as follows

$$\mu_i^{(k)} \equiv E[T_i(\Omega)^k] = (-1)^k \left. \frac{d^k \mathcal{L}_i(z)}{dz^k} \right|_{z=0},$$

and the density function of $T_i(\Omega)$ can be approximated by a numerical inversion of the Laplace-Stieltjes transform $\mathcal{L}_i(z)$ ¹⁵.

We note that the applicability of Eq. (1) relies on the knowledge of (or estimations from experimental data) the (competing) probabilities p_{ij} and the conditioned waiting time transforms $C_{ij}(z)$. In particular, if the process is Markovian (more precisely, a continuous time Markov chain), then all the absolute inter-event times, $\{T_{i \rightarrow j}, j \in A(i)\}$, for any event leaving state i , are exponentially distributed. Mathematically, $T_{i \rightarrow j} \sim \text{Exp}(\lambda_{ij})$, as mentioned in Figs 1b and 2. If this is the case, it is well known that for any $i \in \mathcal{S}$ and $j \in A(i)$, we have

$$p_{ij} = \mathbb{P}(i \rightarrow j) = \frac{\lambda_{ij}}{\sum_{k \in A(i)} \lambda_{ik}}, \quad C_{ij}(z) = E[e^{-zT_{i \rightarrow j}} | i \rightarrow j] = \frac{\sum_{k \in A(i)} \lambda_{ik}}{\sum_{k \in A(i)} \lambda_{ik} + z}.$$

If we substitute these expressions in Eq. (1), one obtains that the transforms $\mathcal{L}_i(z)$ correspond to those of a phase-type distribution (for further details, see Proposition 1.2.2 in¹³), since $T_i(\Omega)$ represents the absorption time for a continuous time Markov chain.

When other distributions other than the exponential are considered, the expressions for the absolute waiting times are not known explicitly. Our aim is to consider alternative distributions to the exponential one for the absolute waiting times $T_{i \rightarrow j}$ with $i \in \mathcal{S}, j \in A(i)$, as represented in Fig. 3, so that we are still able to compute the probabilities p_{ij} and the conditioned transforms $C_{ij}(z)$, and to apply Eq. (1) and its related matrix analytic techniques¹³. In the next section we explore how more general distributions can be considered (under special circumstances), by making use of phase-type distributions. We illustrate our approach making use of several biological case studies in the following sections.

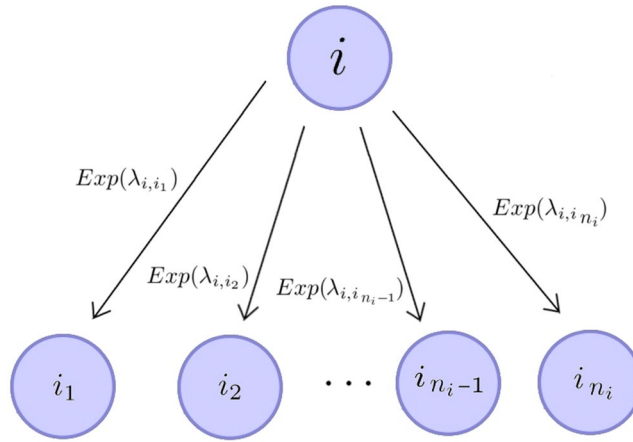


Figure 2. From any state $i \in \mathcal{S}$, the absolute waiting times to any state $j \in A(i) = \{i_1, i_2, \dots, i_{n_i}\}$ are exponentially distributed, $T_{i \rightarrow j} \sim \text{Exp}(\lambda_{ij})$, leading to a process such as the one described in Fig. 1b.

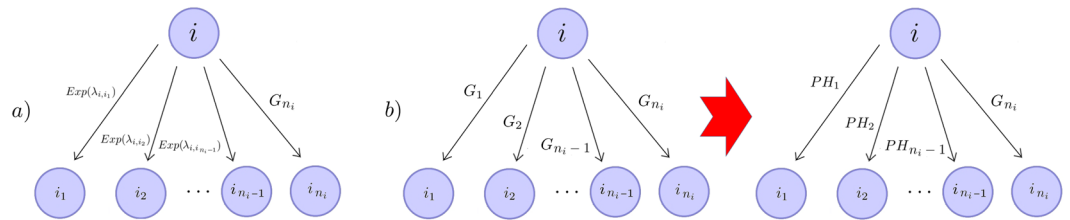


Figure 3. (a) From any given state $i \in \mathcal{S}$, the absolute waiting times to any state $j \in A(i) = \{i_1, i_2, \dots, i_{n_i}\}$ are exponentially distributed, $T_{i \rightarrow j} \sim \text{Exp}(\lambda_{ij})$, except for at most one state i_{n_i} ; (b) If the absolute waiting times are generally distributed, one can approximate all but one of these distributions by phase-type ones, and the methods described in this paper still apply.

Absolute waiting times, transition probabilities and conditioned waiting times for general distributions. Let us assume that for any state in our process, $i \in \mathcal{S}$, with $A(i) = \{i_1, i_2, \dots, i_{n_i}\}$, the absolute inter-event (or waiting) times $T_{i \rightarrow j}$ for $j \in \{i_1, i_2, \dots, i_{n_i-1}\}$, are exponentially distributed, that is, $T_{i \rightarrow j} \sim \text{Exp}(\lambda_{ij})$, and that there is one state $i_{n_i} \in A(i)$ (the last one in the list $A(i)$, without loss of generality) for which this time is generally distributed, $T_{i \rightarrow i_{n_i}} \sim G_{n_i}$ (see Fig. 3a). The distribution G_{n_i} can be described in terms of its Laplace-Stieltjes transform, $\mathcal{L}_{G_{n_i}}(z) = E[e^{-zT_{i \rightarrow i_{n_i}}}]$. Under these conditions, then it is possible to show (see Supplementary Information) that

$$\begin{aligned}
 p_{ii_{n_i}} &= \mathbb{P}(i \rightarrow i_{n_i}) = \mathcal{L}_{G_{n_i}}\left(\sum_{k=i_1}^{i_{n_i-1}} \lambda_{ik}\right), \\
 p_{ij} &= \mathbb{P}(i \rightarrow j) = \frac{\lambda_{ij}}{\sum_{k=i_1}^{i_{n_i-1}} \lambda_{ik}} \left[1 - \mathcal{L}_{G_{n_i}}\left(\sum_{k=i_1}^{i_{n_i-1}} \lambda_{ik}\right)\right], \quad j \in \{i_1, i_2, \dots, i_{n_i-1}\}, \\
 C_{ii_{n_i}}(z) &= \frac{\mathcal{L}_{G_{n_i}}\left(z + \sum_{k=i_1}^{i_{n_i-1}} \lambda_{ik}\right)}{\mathcal{L}_{G_{n_i}}\left(\sum_{k=i_1}^{i_{n_i-1}} \lambda_{ik}\right)}, \\
 C_{ij}(z) &= \frac{\sum_{k=i_1}^{i_{n_i-1}} \lambda_{ik} \left[1 - \mathcal{L}_{G_{n_i}}\left(z + \sum_{k=i_1}^{i_{n_i-1}} \lambda_{ik}\right)\right]}{\sum_{k=i_1}^{i_{n_i-1}} \lambda_{ik} + z \left[1 - \mathcal{L}_{G_{n_i}}\left(\sum_{k=i_1}^{i_{n_i-1}} \lambda_{ik}\right)\right]}, \quad j \in \{i_1, i_2, \dots, i_{n_i-1}\}.
 \end{aligned} \tag{2}$$

Once these quantities are at hand, it is possible to apply Eq. (1) to compute the Laplace-Stieltjes transform of the first passage time $T_i(\Omega)$, for an initial state $i \in \mathcal{S}$, to the given state of interest $\Omega \in \mathcal{S}$.

In Fig. 3b we consider the more interesting case where given any initial state $i \in \mathcal{S}$, all events $i \rightarrow j$ leaving this state have absolute waiting times that are generally distributed, $T_{i \rightarrow i_l} \sim G_l, l \in \{1, 2, \dots, n_i\}$. If this is the case, it is not possible to obtain closed form expressions of the kind given by Eq. (2). However, one can approximate all but one of these distributions (for instance, G_{n_i}) by $n_i - 1$ phase-type distributions $PH_j, j \in \{1, \dots, n_i - 1\}$, leading

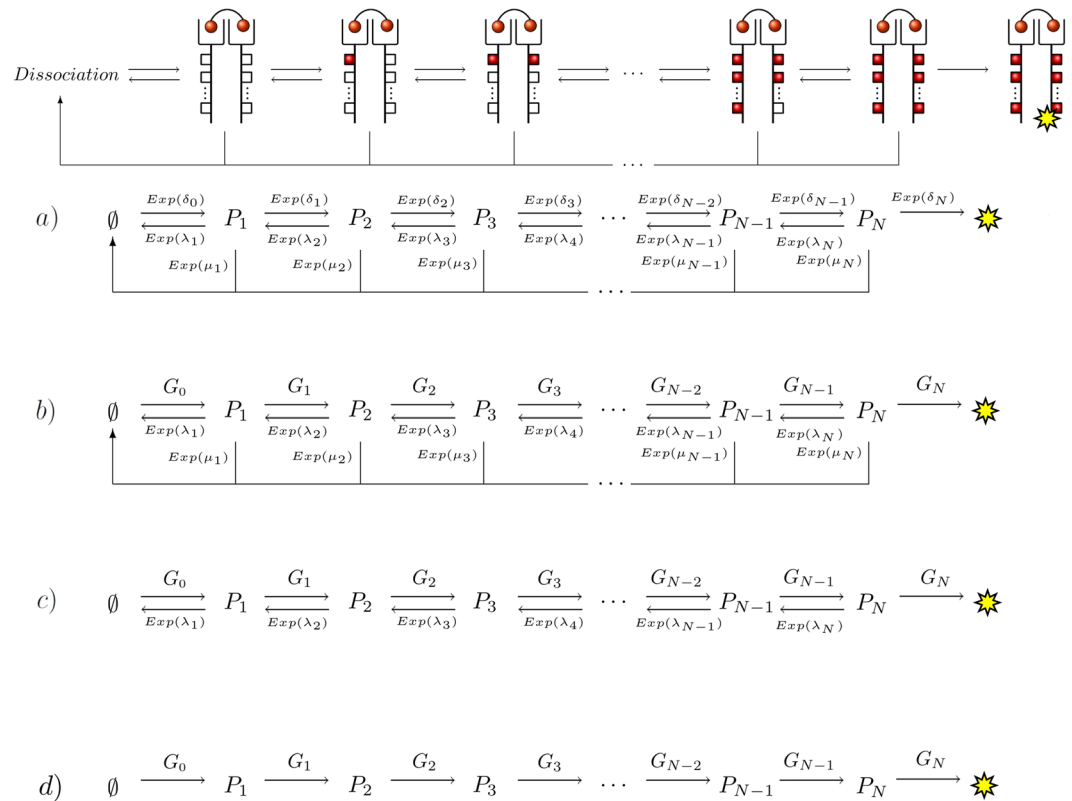


Figure 4. Kinetic proof-reading diagram. **(a)** Markovian model. **(b)** Model with generally distributed phosphorylation events. **(c and d)** Simplified models with some transitions removed. We are interested in studying the distribution of the time to reach the fully phosphorylated state.

to the situation depicted in Fig. 3b right. Thus, by combining exact (general) distributions with arbitrarily accurate approximate phase-type distributions, we can capture the inherent nature of almost any process subject to the semi-Markov restriction. We refer the reader to Supplementary Information, where we have provided all the required mathematical details that allow us to obtain an analogous result to Eq. (2) for the scenario described in Fig. 3b. We note that the approximation of general distributions by phase-type ones is justified, since the family of phase-type (PH) distributions is dense in the family of non-negative continuous distributions (see Theorem 1.2.1 in¹³), and algorithmic approaches have been derived to identify these approximate PH distributions^{13,16,17}. We illustrate this point in case study 2 (Figs 9 and 10, and Supplementary Information). Finally, since the expressions in Eq. (2) depend on the Laplace-Stieltjes transform $\mathcal{L}_{G_{i_j}}(z) = E[e^{-zI_i - i_{i_j}}$], our methods can be implemented even when some of the general distributions in Fig. 3 are discrete or degenerate (*i.e.*, deterministic) waiting time distributions (see Results).

Results

In this section we illustrate our methods by considering three different biological scenarios: receptor-ligand interactions, viral replication and epidemiology. Our first objective is to illustrate the practical implementation of our theoretical methods. Our second objective is to show that the Markovian hypothesis, in which all waiting times are considered exponentially distributed, might not be a good approximation.

Case study 1: kinetic proof-reading. Our first case study considers the signalling dynamics of membrane bound receptors after a ligand-receptor binding and phosphorylation event. In Fig. 4, we show the scenario in which intra-cellular signalling is triggered by a cross-linking event of a bivalent ligand to a receptor dimer¹⁸. This event is followed by a sequence of consecutive and reversible phosphorylation events, which ends when all the phosphorylation sites on the cytoplasmic tail of the receptor are active (or phosphorylated) and the receptor is internalised (thus, initiating the corresponding intra-cellular signalling cascade). If at any time the ligand dissociates from the receptor, the signalling is interrupted *catastrophically*¹⁹ and the process returns to the starting point (pre-binding). This process, known as kinetic proof-reading, provides a robust mechanism for ligand specificity and has been widely studied in the literature¹, even from a stochastic (Markovian) perspective²⁰.

In Fig. 4 we consider four possible scenarios for the process \mathcal{X} that involves states $\{0, 1, 2, \dots, N\} \cup \{\Omega\}$, where 0 represents the pre-binding state (\emptyset) and Ω is the final (and fully) phosphorylated state, represented with a yellow star in Fig. 4. In Fig. 4a we summarise the Markovian view as introduced in ref.²⁰. In this reference, all the transitions are Markovian and, thus, inter-event times are assumed to be exponentially distributed. In Fig. 4b we

relax this assumption by incorporating generally distributed absolute waiting times for the phosphorylation events, so that the arguments of the Materials and Methods section still apply. In particular, we recognise that the event that allows the phosphorylation of a new site on the intra-cellular tail of the receptor requires (at least) the diffusion and binding to this tail of adaptor and scaffold proteins²¹, so that this waiting time could have a more complex form than exponential²². In Fig. 4c, we consider the case of a strong agonist (a ligand with a large affinity to the receptor, so that phosphorylation events occur in times short enough to assume that ligand dissociation cannot occur). Finally, in Fig. 4d, we assume that phosphorylation events are not reversible. This could represent a scenario where selected intra-cellular phosphatases have been deleted, so that dephosphorylation events cannot take place.

In our framework, these four scenarios can be summarised by providing the matrix $A(z)$ in Eq. (1). We recall that if an absolute waiting time is exponentially distributed, $T_{i \rightarrow j} \sim \text{Exp}(\lambda_{ij})$, its Laplace-Stieltjes transform is given by

$$\mathcal{L}_{\text{Exp}(\lambda_{ij})}(z) = E[e^{-zT_{i \rightarrow j}}] = \frac{\lambda_{ij}}{\lambda_{ij} + z}. \tag{3}$$

We denote by $T_{\text{Exp}(\lambda_j)}$, $T_{\text{Exp}(\mu_j)}$, $T_{\text{Exp}(\delta_j)}$ and T_{G_j} , the exponentially distributed and generally distributed absolute waiting times shown in Fig. 4. By providing suitable expressions for p_{ij} and $C_{ij}(z)$ in Eq. (1), making use of the arguments in the Supplementary Information and Eq. (2), we can express Fig. 4 in matrix form. We note that, in all cases, the matrix structure captures the arrow structure of the diagrams shown in Fig. 4. Each of these matrices is given below.

$$A^a(z) = \begin{pmatrix} 1 & -\frac{\delta_0}{\delta_0 + z} & 0 & 0 & 0 & \dots & 0 \\ -\frac{\mu_1 + \lambda_1}{\mu_1 + \lambda_1 + \delta_1 + z} & 1 & -\frac{\delta_1}{\mu_1 + \lambda_1 + \delta_1 + z} & 0 & 0 & \dots & 0 \\ -\frac{\mu_2}{\mu_2 + \lambda_2 + \delta_2 + z} & -\frac{\lambda_2}{\mu_2 + \lambda_2 + \delta_2 + z} & 1 & -\frac{\delta_2}{\mu_2 + \lambda_2 + \delta_2 + z} & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\frac{\mu_N}{\mu_N + \lambda_N + \delta_N + z} & 0 & 0 & \dots & -\frac{\lambda_N}{\mu_N + \lambda_N + \delta_N + z} & 1 & -\frac{\delta_N}{\mu_N + \lambda_N + \delta_N + z} \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}$$

$$A^b(z) = \begin{pmatrix} 1 & -\mathcal{L}_{G_0}(z) & 0 & \dots & 0 \\ -\frac{\mu_1 + \lambda_1}{\mu_1 + \lambda_1 + z}(1 - \mathcal{L}_{G_1}(\mu_1 + \lambda_1 + z)) & 1 & -\mathcal{L}_{G_1}(\mu_1 + \lambda_1 + z) & \dots & 0 \\ -\frac{\mu_2}{\mu_2 + \lambda_2 + z}(1 - \mathcal{L}_{G_2}(\mu_2 + \lambda_2 + z)) & -\frac{\lambda_2}{\mu_2 + \lambda_2 + z}(1 - \mathcal{L}_{G_2}(\mu_2 + \lambda_2 + z)) & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -\frac{\mu_N}{\mu_N + \lambda_N + z}(1 - \mathcal{L}_{G_N}(\mu_N + \lambda_N + z)) & 0 & \dots & 1 - \mathcal{L}_{G_N}(\mu_N + \lambda_N + z) \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

We note that the case in Fig. 4a can be obtained from the latter matrix by setting the distributions $G_j \equiv \text{Exp}(\delta_j)$. In this case, $\mathcal{L}_{G_j}(z) = \frac{\delta_j}{\delta_j + z}$ and the resulting $A^b(z)$ is exactly the matrix obtained for case a).

$$A^c(z) = \begin{pmatrix} 1 & -\mathcal{L}_{G_0}(z) & 0 & 0 & \dots & 0 \\ -\frac{\lambda_1}{\lambda_1 + z}(1 - \mathcal{L}_{G_1}(\lambda_1 + z)) & 1 & -\mathcal{L}_{G_1}(\lambda_1 + z) & 0 & \dots & 0 \\ 0 & -\frac{\lambda_2}{\lambda_2 + z}(1 - \mathcal{L}_{G_2}(\lambda_2 + z)) & 1 & -\mathcal{L}_{G_2}(\lambda_2 + z) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\frac{\lambda_N}{\lambda_N + z}(1 - \mathcal{L}_{G_N}(\lambda_N + z)) & 1 & -\mathcal{L}_{G_N}(\lambda_N + z) \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}$$

Again, this case can be seen as a particular case of b) when $\mu_j = 0$ for $1 \leq j \leq N$.

$$A^d(z) = \begin{pmatrix} 1 & -\mathcal{L}_{G_0}(z) & 0 & 0 & \dots & 0 \\ 0 & 1 & -\mathcal{L}_{G_1}(z) & 0 & \dots & 0 \\ 0 & 0 & 1 & -\mathcal{L}_{G_2}(z) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1 & -\mathcal{L}_{G_N}(z) \\ 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}$$

As the latter matrix is bidiagonal, it is straightforward to see that in this case $\mathcal{L}_i(z) = E[e^{-zT_i(\Omega)}] = (A^{-1}(z))_{i,N+1} = \prod_{k=i}^N \mathcal{L}_{G_k}(z)$, which is the well known result for a pure birth process²³. That is, the time to reach state Ω , starting from state i , is the sum of the independent and generally distributed times $T_i(\Omega) = T_{G_i} + T_{G_{i+1}} + \dots + T_{G_N}$, so that $\mathcal{L}_i(z)$ is the product of their corresponding transforms. Moreover, if $\mathcal{L}_{G_k}(z) = \lambda/(\lambda + z)$ for all k (that is, if $G_k \equiv \text{Exp}(\lambda)$ for some $\lambda > 0$), $\mathcal{L}_i(z)$ is the Laplace-Stieltjes transform of an Erlang distribution, as one would expect.

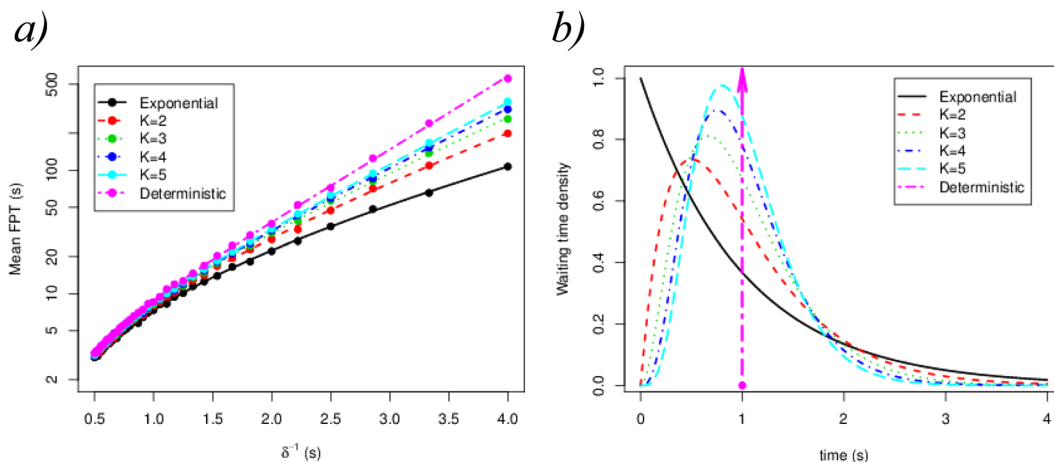


Figure 5. (a) Mean FPT $E[T_0(\Omega)]$ from pre-binding (state 0) until signalling (state Ω) for the model described in Fig. 4b. The probability distributions considered for the absolute waiting time are shown in panel (b). Every waiting time distribution, G , in panel (b) has the same mean, $\delta^{-1} = 1s$, but different shape. The value K refers to the shape parameter of the gamma distribution ($K = 1$ corresponds to the exponential distribution). We have considered $N = 4$, $\lambda_k = 0.2s^{-1}$ and $\mu_k = 0.1s^{-1}$ for all k in Fig. 4, for illustrative purposes.

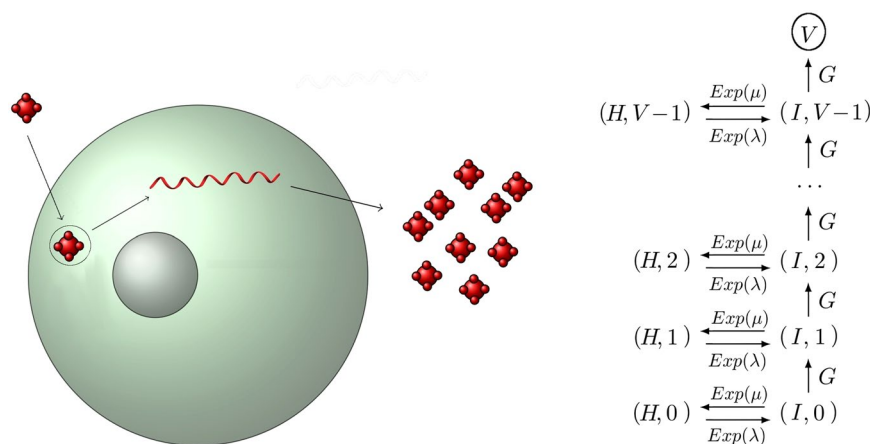


Figure 6. Viral budding for a cell that switches between healthy (H) and infected (I) states.

The transforms $\mathcal{L}_i(z)$ can be obtained by solving Eq. (1) for the matrices $\mathbf{A}(z)$ above. For the general case b), and if we define $\Delta_j = \lambda_j + \mu_j$, a recursive solution is given as follows:

$$\begin{aligned}
 h_0(z) &= 1. \\
 h_1(z) &= 1 - \frac{\Delta_1}{\Delta_1 + z} [1 - \mathcal{L}_{G_1}(\Delta_1 + z)] \mathcal{L}_{G_0}(z). \\
 h_2(z) &= 1 - [1 - \mathcal{L}_{G_2}(\Delta_2 + z)] h_1(z)^{-1} \mathcal{L}_{G_1}(\Delta_1 + z) \left(\frac{\mu_2 \mathcal{L}_{G_0}(z) + \lambda_2}{\Delta_2 + z} \right). \\
 h_3(z) &= 1 - [1 - \mathcal{L}_{G_3}(\Delta_3 + z)] h_2(z)^{-1} \mathcal{L}_{G_2}(\Delta_2 + z) \left(\frac{\mu_3 \mathcal{L}_{G_0}(z) h_1(z)^{-1} \mathcal{L}_{G_1}(\Delta_1 + z) + \lambda_3}{\Delta_3 + z} \right).
 \end{aligned}$$

For $k = 4, \dots, N$:

$$\begin{aligned}
 h_k(z) &= 1 - [1 - \mathcal{L}_{G_k}(\Delta_k + z)] h_{k-1}(z)^{-1} \mathcal{L}_{G_{k-1}}(\Delta_{k-1} \\
 &\quad + z) \left(\frac{\mu_k \mathcal{L}_{G_0}(z) h_1(z)^{-1} \mathcal{L}_{G_1}(\Delta_1 + z) \dots h_{k-2}(z)^{-1} \mathcal{L}_{G_{k-2}}(\Delta_{k-2} + z) + \lambda_k}{\Delta_k + z} \right). \\
 \mathcal{L}_N(z) &= h_N(z)^{-1} \mathcal{L}_{G_N}(\Delta_N + z).
 \end{aligned}$$

For $k = N - 1, N - 2, \dots, 0$:

$$\mathcal{L}_k(z) = h_k(z)^{-1} \mathcal{L}_{G_k}(\Delta_k + z) \mathcal{L}_{k+1}(z). \tag{4}$$

This recursive scheme is also valid for *a*) and *c*), since these are particular cases of case *b*). We note that Eq. (4) for the scenario considered in Fig. 4d would be given by, for any initial state $0 \leq i \leq N$, the following expression

$$\mathcal{L}_i(z) = \prod_{k=i}^N \mathcal{L}_{G_k}(z) \Rightarrow E[T_0(\Omega)] = \sum_{k=0}^N E[T_{k \rightarrow k+1}].$$

Namely, the mean time to reach state Ω (mean first passage time (FPT)) starting at 0, $E[T_0(\Omega)]$, is the sum of the mean times spent in every transition of the process. This result is also what one would expect under the assumption that every transition can be described by an exponential rate δ_k , where we obtain

$$E[T_0(\Omega)] = \sum_{k=0}^N \frac{1}{\delta_k},$$

which only depends on the mean times of each transition, δ_k^{-1} . However, the mean of each transition event is not enough when analysing the mean FPT to state Ω if general distributions are considered for scenarios more complex than Fig. 4d, such as those described by Fig. 4b,c. In particular, when studying the mean FPT to reach state Ω from the initial state 0 (see Fig. 5), we set the mean time for a phosphorylation event to be the same, $E[T_{G_k}] = \delta^{-1}$ independently of k and we set the distribution G ($G_k \equiv G$ for all k) and consider the case given by Fig. 4a and that of Fig. 4b (that is, $G \equiv \text{Exp}(\delta)$ or other general distributions). In particular, we use an exponential distribution (corresponding to Fig. 4a) and gamma distributions with shape $K = 2, \dots, 5$ and a deterministic distribution with time δ^{-1} (see densities in Fig. 5b). The symbols in Fig. 5a stand for the numerical simulations of the diagram in Fig. 4 and the lines for the theoretical prediction provided by Eq. (4).

Figure 5 clearly shows that the shape of the distribution G can dramatically change the estimated mean signalling time $E[T_0(\Omega)]$ (note that the vertical axis is a logarithmic scale), even when $E[T_G] = \delta^{-1}$ is the same. We note that an important result derived from our approach, and reported in the Supplementary Information, is that from the empirical absolute waiting time distributions, we can relate the conditioned times, $T_{i \rightarrow j} | i \rightarrow j$ to the absolute ones, $T_{i \rightarrow j}$. Finally, we stress the fact that these general results can be applied to the generic case (Fig. 4b), or in the presence of an agonist ligand (Fig. 4c) or in the absence of phosphatases (Fig. 4d).

Case study 2: virion production. Our next case study involves the continuous release of virions from infected cells. Our model consists of a cell that can be in two states: *healthy* (H) or *infected* (I). In the latter, the cell releases one virion at a time after generally distributed times, T_G , with common general distribution G , and Laplace-Stieltjes transform $\mathcal{L}_G(z) = E[e^{-zT_G}]$. The transition between the infected and healthy states is assumed to be exponentially distributed (with rates λ and μ), as shown in Fig. 6. A natural question is to obtain the distribution of the total time, $T_{(H,0)}(V)$, it takes to produce V virions, starting from state $(H, 0)$, that is, at time $t = 0$ the cell is in the healthy state H and no virions have been released yet.

Using the definitions provided in the Materials and Methods section, and by ordering the space of states as follows

$$(H, 0) \prec (I, 0) \prec (H, 1) \prec (I, 1) \prec \dots \prec (H, V-1) \prec (I, V-1) \prec V,$$

we can write the matrix

$$\mathbf{A}(z) = \begin{pmatrix} 1 & -\frac{\lambda}{\lambda+z} & 0 & 0 & 0 & 0 & \dots & 0 \\ -\frac{\mu}{\mu+z}(1 - \mathcal{L}_G(\mu+z)) & 1 & 0 & -\mathcal{L}_G(\mu+z) & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & -\frac{\lambda}{\lambda+z} & 0 & 0 & \dots & 0 \\ 0 & 0 & -\frac{\mu}{\mu+z}(1 - \mathcal{L}_G(\mu+z)) & 1 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & \dots & \dots & 1 & -\mathcal{L}_G(\mu+z) \end{pmatrix}.$$

Following the same procedure as in the previous case study, we find that for the initial state $(H, 0)$,

$$\mathcal{L}_{(H,0)}(z) = E[e^{-zT_{(H,0)}(V)}] = \frac{\lambda}{\lambda+z} \left(\frac{(\lambda+z)(\mu+z)\mathcal{L}_G(\mu+z)}{(\lambda\mu\mathcal{L}_G(\mu+z) + z(\lambda+\mu+z))} \right)^V,$$

and thus, the mean time to produce V virions is given by

$$E[T_{(H,0)}(V)] = -\frac{d\mathcal{L}_{(H,0)}(z)}{dz} \Big|_{z=0} = \frac{\mathcal{L}_G(\mu)(\mu - V(\lambda + \mu)) + V(\lambda + \mu)}{\lambda\mu\mathcal{L}_G(\mu)} = \frac{1}{\lambda} + V \left(\frac{1}{\mu} + \frac{1}{\lambda} \right) [\mathcal{L}_G(\mu)^{-1} - 1]. \tag{5}$$

This expression can be easily understood: the mean time to produce V virions can be split as the sum of the mean time to get the cell infected ($1/\lambda$) and the mean time per new virion production, $\Delta T = \left(\frac{1}{\mu} + \frac{1}{\lambda} \right) [\mathcal{L}_G(\mu)^{-1} - 1]$, times the number of required virions, V . We note that ΔT depends on $\mathcal{L}_G(\mu)$.

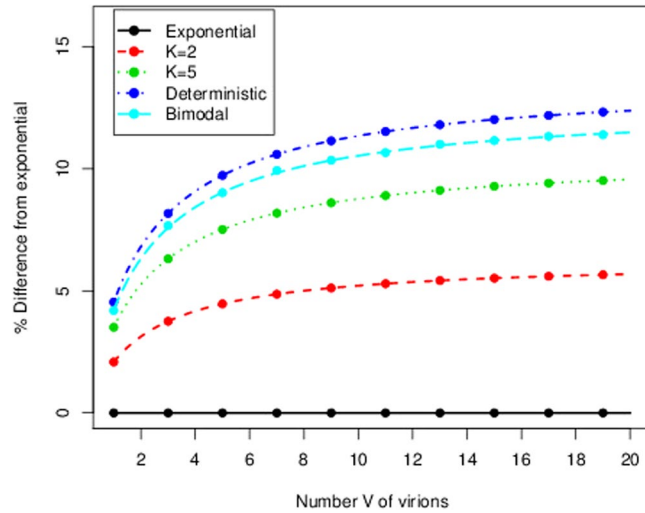


Figure 7. Percentage difference between the mean FPT $E[T_{(H,0)}(V)]$ under the Markovian hypothesis and under general distributions G for each budding event (see Fig. 6). A common mean $\tau = 0.5h$ has been chosen, with $\mu = \lambda = 0.5h^{-1}$. The label *bimodal* refers to a mixture $qN(\tau_1, \sigma^2) + (1 - q)N(\tau_2, \sigma^2)$ of Gaussian distributions with values $\sigma = 0.05, q = 0.26, \tau_1 = 0.26h$ and $\tau_2 = 0.58h$, so that $\tau = q\tau_1 + (1 - q)\tau_2 = 0.5h$. Dots show the results from stochastic simulations and lines show the theoretical results obtained from our matrix analytic approach.

Thus, the shape of the general waiting time distribution G changes the prediction of the mean time to produce V virions. For a purely Markovian process, $G \equiv \text{Exp}\left(\frac{1}{\tau}\right)$, where τ denotes the mean virion production time, we have

$$\mathcal{L}_G(\mu)^{-1} - 1 = \mu\tau, \tag{6}$$

which only depends on $E[T_G] = \tau$. However, for more general distributions, $\mathcal{L}_G(\mu)$ would depend also on all the moments (or cumulants, such as the variance, σ^2) of the general distribution G , since

$$\mathcal{L}_G(\mu)^{-1} - 1 = \mu\tau + \frac{1}{2}\mu^2(\tau^2 - \sigma^2) + \dots(\text{higher moments}).$$

These differences are highlighted in the numerical simulations which follow.

As we discussed above, the time to produce V virions is the sum of the waiting time until the first infection plus V times the time in a production-recovery-reinfection cycle. In Fig. 7 we show the percentage difference between the mean FPT $E[T_{(H,0)}(V)]$ under the Markovian hypothesis ($G \equiv \text{Exp}\left(\frac{1}{\tau}\right)$ in Fig. 6) and under generally distributed absolute waiting times for the budding event. The bimodal distribution is a mixture of two Gaussian distributions centered at times τ_1 and τ_2 , inspired by the experimental observations discussed in ref.⁴. We have chosen the parameters of the general distributions so that they have the same mean, τ , as the exponential case.

Inspection of Fig. 7 reveals two important features. First, the numerical simulations follow the theoretical predictions for any choice of distribution G . Secondly, and as we have already discussed in the Results section, the shape of the distribution G , and not only its mean, is an important factor to estimate the mean time to produce V virions. For the parameters used in Fig. 7, these differences can be as large as 12%. In fact, these differences will be larger or smaller depending on the form of the Laplace-Stieljes transform of the absolute waiting time distribution considered, and whether this form can or cannot be suitably captured by its first few moments (for instance, if the distribution is not monotonic). The largest differences are obtained when $G \equiv \text{Deterministic} \equiv \tau$ is considered. This fact, which would not be true in general for an arbitrary model, is true for the particular model described in Fig. 6. Note that, from Jensen’s inequality, for any arbitrary density function of a random variable X , the following inequality is satisfied

$$\exp(-zE[X]) \leq E[\exp(-zX)].$$

If the term on the right is the definition of the Laplace transform of the general distribution G , and considering that the general distribution has mean τ , we find

$$e^{-z\tau} \leq \mathcal{L}_G(z)$$

Interestingly, $e^{-z\tau}$ is the Laplace transform of the deterministic distribution. Hence, as the mean time to produce V virions is proportional to $\mathcal{L}_G(\mu)^{-1} - 1$, by Eq. (5), this proves that the deterministic distribution constitutes an upper bound for $E[T_{(H,0)}(V)]$ in Fig. 7.

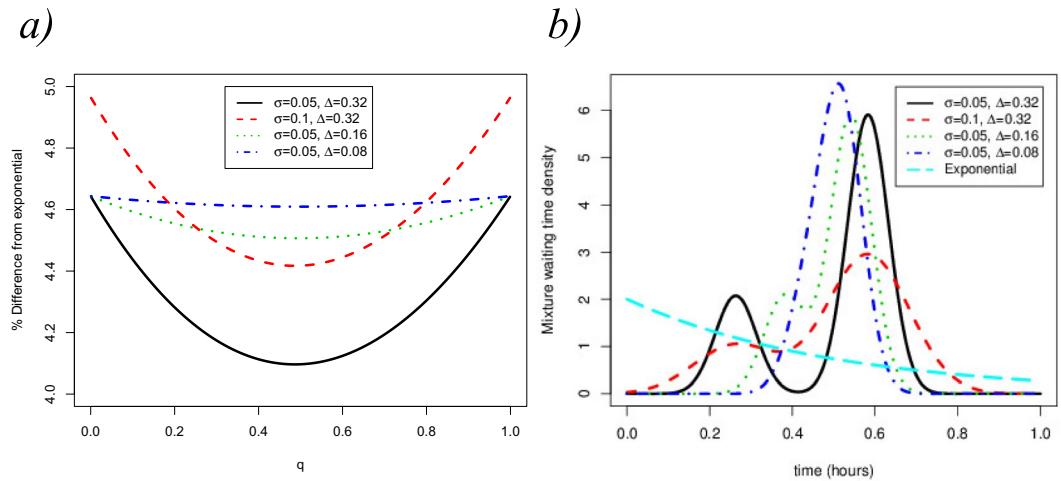


Figure 8. (a) Relative difference in the mean time to produce $V = 1$ virions between the Markovian hypothesis ($G \equiv \text{Exp}(\frac{1}{\tau})$) and the choice of G as a mixture of Gaussian distributions with different shifts, Δ , and variances, σ^2 . We note that values of τ_1 and τ_2 are adjusted according to Eq. (7), so that $E[T_G] = \tau$. (b) Plot of the distribution G for $q = 0.26$.

We further illustrate the impact of the particular choice of the distribution G in the first passage time under analysis by plotting (Fig. 8) the theoretical relative difference between the Markovian hypothesis ($G \equiv \text{Exp}(\frac{1}{\tau})$) and the choice of G as a mixture of Gaussian distributions of the form

$$T_G \sim qN(\tau_1, \sigma^2) + (1 - q)N(\tau_2, \sigma^2)$$

where τ is the mean time of the exponential distribution as discussed in Eq. (6) and we vary, keeping τ fixed, as follows

$$\tau_2 - \tau_1 = \Delta, \quad q\tau_1 + (1 - q)\tau_2 = \tau. \tag{7}$$

The parameter σ accounts for the common dispersion around each peak of the bimodal distribution and Δ for the shift between the peaks. From Fig. 8 we can see that both parameters Δ and σ have an impact on the % difference. Furthermore, Fig. 8 suggests that a minimal difference is obtained for $q = 0.5$, which represents the mixture of two Gaussian distributions with equal probability. It is worth noting that using an exponential distribution always underestimates the mean FPT.

An additional strength of our approach is that we can extend these results to the case of phase-type (PH) waiting times. In particular, we can study a general scenario containing, for some initial state i , n_i general waiting time distributions, by approximating the process as follows: one such waiting time would be generally distributed and the remaining $n_i - 1$ waiting times would be PH distributions, with these PH distributions approximating the $n_i - 1$ general ones with a desired degree of accuracy (see Fig. 3). This is justified by the fact that any non-negative continuous distribution can be approximated by a PH distribution¹³, [Theorem 1.2.1]. To test the accuracy of this approximation within our framework, we have simulated this case study, but replacing the recovery event (originally an exponential distribution with rate μ) by a log-normal distribution with the same mean μ^{-1} and standard deviation $\sigma = 0.5$. We note that the use of a phase-type distribution as an approximation is equivalent to adding a number of ghost states in the system, see Fig. 9.

Figure 10a shows a (black) histogram obtained by drawing 10^4 sampled values from the log-normal distribution against the density curve (red) of the phase-type distribution $PH(\alpha, \mathbf{T})$ defined by

$$\alpha = (1, 0, 0),$$

$$\mathbf{T} = \begin{pmatrix} -1.5186 & 1.5155 & 0 \\ 0 & -1.5971 & 1.5871 \\ 0 & 0 & -1.5235 \end{pmatrix},$$

which has been obtained using the PhaseType R package¹⁶. Figure 10a allows us to compare a three-dimensional PH (approximation) probability density (solid red line) to a histogram (black) of the corresponding log-normal distribution. The mean time, obtained by stochastic simulations, to produce V virions combining a mixture distribution for G , as defined above (budding events), and a log-normal one for recovery events (instead of $\text{Exp}(\mu)$) is given by the filled black circles in Fig. 10b. Different approximations, where these two distributions are replaced either by an exponential one or by a PH distribution, yield the theoretically computed curves in Fig. 10b. It is to be expected that higher dimensional PH approximations of the log-normal would result in better approximations. Nevertheless, as shown in Fig. 10b, the black solid line is the theoretical prediction arising when the log-normal recovery time distribution is substituted by this PH distribution (see Supplementary Information for further details), while keeping the general bimodal distribution G for the budding events. The prediction of the theory

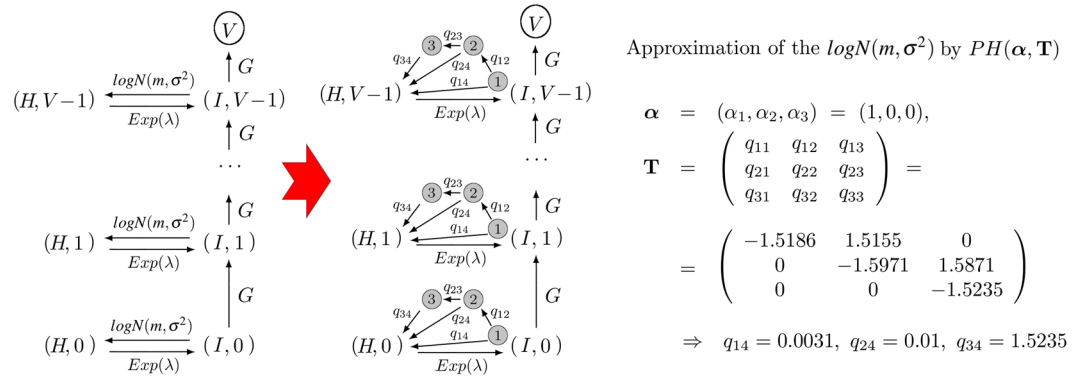


Figure 9. Process in Fig. 6 where recovery of the cell occurs after a $\log N(m, \sigma^2)$ distributed time, instead of $Exp(\mu)$. We consider $m = 0.57$ h, $\sigma = 0.5$ h so that $1/\mu = 2$ h. We approximate the log-normal recovery time with a phase-type distribution, $PH(\alpha, \mathbf{T})$, using three phases.

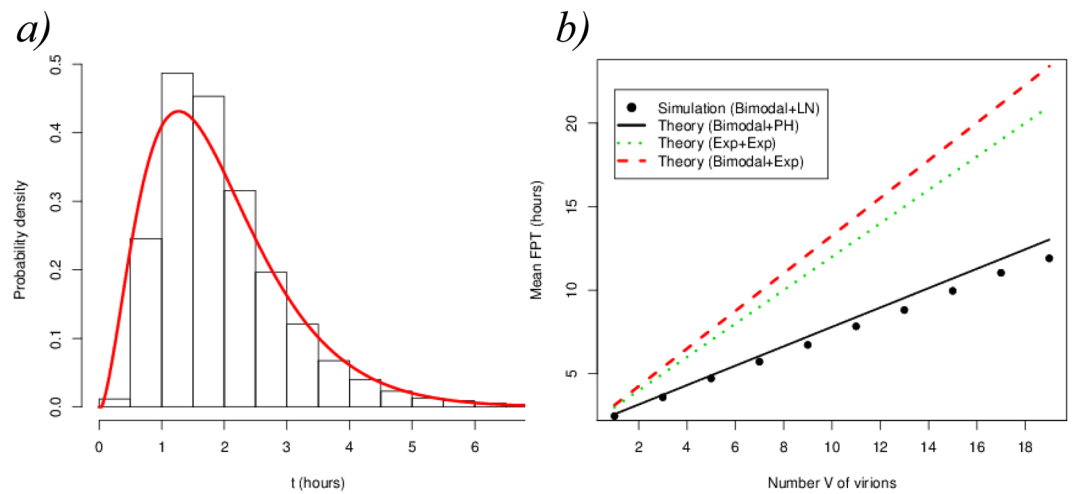


Figure 10. (a) Bars: histogram of the log-normal distribution $\log N(m, \sigma^2)$ with $m = 0.57$ h, $\sigma = 0.5$ h so that $1/\mu = 2$ h. Red solid line: probability density of the three-dimensional $PH(\alpha, \mathbf{T})$ approximation of the log-normal distribution. (b) Black circles: simulation of case study 2 with a mixture of Gaussian distributions for budding events and a log-normal one for the recovery events. The black solid line represents the theoretical prediction when replacing the log-normal by a PH distribution in panel (a). The red dashed line represents the theoretical prediction using a mixture of Gaussian distributions for the budding event and an exponential one for the recovery transition. The green dotted line represents the complete Markovian hypothesis.

captures accurately the simulations and shows the large overestimations arising from different approximations that make use of exponentially distributed waiting times. We note that only the bimodal plus PH theoretical approximation can account for the simulation results that correspond to the bimodal plus log-normal *real* system.

Case study 3: interventions in epidemiology. Our last case deals with an epidemiological problem. We consider a medical doctor working in a hospital ward in which some bacteria have been detected to be spreading among patients and healthcare workers (HCWs). When this outbreak is detected ($t = 0$), this doctor is in a healthy (*susceptible*) state, and she/he can become infected due to treating the (potentially infected) patients at the ward during this outbreak. We consider the following assumptions:

- (i) The outbreak finishes in an average time $\frac{1}{\delta}$, which we consider to be exponentially distributed, $Exp(\delta)$.
- (ii) During this outbreak, the doctor can get infected after some average time $\frac{1}{\beta}$, which we consider to be exponentially distributed, $Exp(\beta)$.
- (iii) Once the doctor is infected, she/he can naturally recover after a mean time $\frac{1}{\gamma}$, and we consider this time to be exponentially distributed, $Exp(\gamma)$.

The Markovian hypotheses above are the usual ones when studying this type of processes from a stochastic perspective. However, we also incorporate generally distributed times to describe the following hospital

- surveillance (i.e., screening) policies:
- (iv) Since the detection of the outbreak ($t=0$), a new policy has been implemented so that every HCW is screened, every B days, to check if she/he is infected. If the HCW is found to be infected, she/he is sent off-duty.
 - (v) Once the HCW is sent off-duty, she/he is given some treatment so that her/his recovery occurs after a time that is exponentially distributed, $Exp(\tilde{\gamma})$, with $\tilde{\gamma} \geq \gamma$, and she/he is screened every M days to check for a complete recovery.

We are interested in studying the process until the outbreak ends (if the HCW under surveillance does not become infected), or until the recovery of this individual is completed (if she/he suffers the infection). In order to analyse this process, we consider the diagram in Fig. 11. We study the process until it reaches the state Δ , which represents either the end of the outbreak at the hospital ward without the HCW under study becoming infected, or the recovery of this HCW if she/he suffers the infection.

The general distribution $G_{I \rightarrow OD}$ describes the time $T_{I \rightarrow OD}$ that it takes to detect the infection of the HCW, by means of the screening policy (SP) implemented in the hospital ward, defined in (iv). Thus, once the HCW is infected, the time $T_{I \rightarrow OD}$ for the infection of this HCW to be detected depends on how close to the next screening event this infection event actually occurred. We note that the time to infection of the individual, conditioned on this infection occurring, $T_{S \rightarrow I} | S \rightarrow I \sim Exp(\beta + \delta)$. This is true from the properties of the exponential distribution, thus the distribution of $T_{I \rightarrow OD}$ can be described in terms of its density function as follows

$$f_{T_{I \rightarrow OD}}(t) = \sum_{k=0}^{+\infty} f_{T_{S \rightarrow I} | S \rightarrow I}((k+1)B - t) = \sum_{k=0}^{+\infty} (\beta + \delta) e^{-(\beta + \delta)((k+1)B - t)}$$

$$= (\beta + \delta) e^{(\beta + \delta)(t - B)} \sum_{k=0}^{+\infty} e^{-(\beta + \delta)kB} = \frac{(\beta + \delta) e^{(\beta + \delta)(t - B)}}{1 - e^{-(\beta + \delta)B}}, \quad 0 < t < B.$$

Thus, the Laplace-Stieltjes transform of the time $T_{I \rightarrow OD} \sim G_{I \rightarrow OD}$ is given by

$$\mathcal{L}_{G_{I \rightarrow OD}}(z) = \int_0^B e^{-zt} dt \frac{(\beta + \delta) e^{(\beta + \delta)(t - B)}}{1 - e^{-(\beta + \delta)B}} = \frac{(\beta + \delta) e^{-(\beta + \delta)B} [e^{(\beta + \delta - z)B} - 1]}{(\beta + \delta - z) [1 - e^{-(\beta + \delta)B}]}$$

On the other hand, the general distribution $G_{OD \rightarrow \Delta}$ describes the time $T_{OD \rightarrow \Delta}$ that it takes to detect the recovery of the off-duty HCW. That is, the sum of the times for the individual to recover and then for this recovery to be detected by the screening policy in (v). From the SP (v), it is clear that $T_{OD \rightarrow \Delta}$ is a discrete random variable with the following probability mass function

$$\mathbb{P}(T_{OD \rightarrow \Delta} = kM) = \mathbb{P}((k-1)M < Exp(\tilde{\gamma}) < kM) = e^{-\tilde{\gamma}(k-1)M} - e^{-\tilde{\gamma}kM}, \quad k \in \{1, 2, 3, \dots\}.$$

Hence, the Laplace-Stieltjes transform of this distribution is given by

$$\mathcal{L}_{G_{OD \rightarrow \Delta}}(z) = \sum_{k=1}^{+\infty} e^{-zkM} (e^{-\tilde{\gamma}(k-1)M} - e^{-\tilde{\gamma}kM}) = \frac{e^{-(z + \tilde{\gamma})M} (e^{\tilde{\gamma}M} - 1)}{1 - e^{-(z + \tilde{\gamma})M}}$$

We now order the states as follows: $S \prec I \prec OD \prec \Delta$, so that the matrix $A(z)$ is given by

$$A(z) = \begin{pmatrix} 1 - \frac{\beta}{\beta + \delta + z} & 0 & -\frac{\delta}{\beta + \delta + z} \\ 0 & 1 & -\mathcal{L}_{G_{I \rightarrow OD}}(z + \gamma) - \frac{\gamma}{\gamma + z} (1 - \mathcal{L}_{G_{I \rightarrow OD}}(\gamma + z)) \\ 0 & 0 & 1 & -\mathcal{L}_{G_{OD \rightarrow \Delta}}(z) \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{8}$$

We conclude then, that the Laplace-Stieltjes transform of the time $T_S(\Delta)$ to reach state Δ , from the initial state S , $\mathcal{L}_S(z) = E[e^{-zT_S(\Delta)}]$, is given by

$$\mathcal{L}_S(z) = \frac{\beta(\beta + \delta) (e^{\tilde{\gamma}M} - 1) (e^{B(\beta - \gamma + \delta - z)} - 1) e^{B(-\beta - \delta) + M(-\tilde{\gamma} - z)}}{(1 - e^{B(-\beta - \delta)}) (1 - e^{M(-\tilde{\gamma} - z)}) (\beta + \delta + z) (\beta - \gamma + \delta - z)} + \frac{\beta \gamma \left(1 - \frac{(\beta + \delta) e^{B(-\beta - \delta)} (e^{B(\beta - \gamma + \delta - z)} - 1)}{(1 - e^{B(-\beta - \delta)}) (\beta - \gamma + \delta - z)} \right)}{(\gamma + z) (\beta + \delta + z)} + \frac{\delta}{\beta + \delta + z}, \tag{9}$$

and the mean of the distribution is given by

$$E[T_S] = \frac{1}{\gamma(\beta + \delta) [e^{\tilde{\gamma}M} - 1] (\beta - \gamma + \delta) (e^{B(\beta + \delta)} - 1)} e^{-B\gamma} (-(\beta + \gamma) (\beta - \gamma + \delta) e^{B(\beta + \gamma + \delta)} + \beta(\beta + \delta) e^{B(\beta + \delta)} - \gamma e^{B\gamma} (\gamma - \delta) + \beta(\beta + \delta) (\gamma M - 1) e^{B(\beta + \delta) + \tilde{\gamma}M} + (\beta + \gamma) (\beta - \gamma + \delta) e^{B(\beta + \gamma + \delta) + \tilde{\gamma}M} + \gamma e^{B\gamma + \tilde{\gamma}M} (\gamma - \delta - \beta M(\beta + \delta))). \tag{10}$$

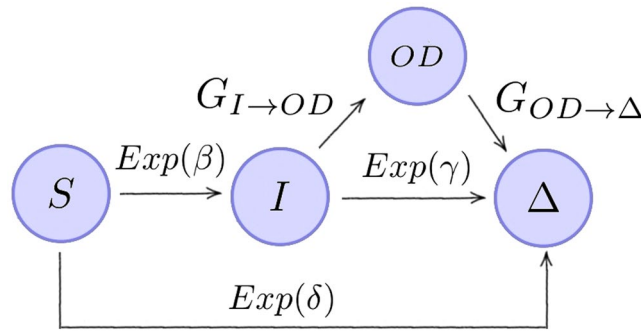


Figure 11. Single HCW epidemic model. General distributions $G_{I \rightarrow OD}$ and $G_{OD \rightarrow \Delta}$ correspond to screening policies (iv) and (v) described in the main text.

In order to show how this approach can also be applied to compute other quantities of interest, and not only first passage times, we now calculate the following probabilities: the probability of the medical doctor not becoming infected, $p_S(\overline{INF})$, suffering the infection and being sent off-duty, $p_S(INF \cap DET)$, and suffering the infection but not being detected $p_S(INF \cap \overline{DET})$. We know that

$$p_S(\overline{INF}) + p_S(INF \cap DET) + p_S(INF \cap \overline{DET}) = 1,$$

with

$$p_S(\overline{INF}) = \frac{\delta}{\beta + \delta},$$

$$p_S(INF) = p_S(INF \cap DET) + p_S(INF \cap \overline{DET}) = \frac{\beta}{\beta + \delta}.$$

To compute the probabilities of being infected and detected, or infected but not detected, we can use the following system of equations obtained from a first step argument

$$\begin{aligned} p_S(INF \cap DET) &= p_{SI}p_I(INF \cap DET) + p_{S\Delta}p_{\Delta}(INF \cap DET), \\ p_I(INF \cap DET) &= p_{I,OD}p_{OD}(INF \cap DET) + p_{I\Delta}p_{\Delta}(INF \cap DET), \end{aligned}$$

where $p_{ij} = \mathbb{P}(i \rightarrow j)$ can be obtained from the arguments described in the Materials and Methods section and the Supplementary Information, and boundary conditions $p_{\Delta}(INF \cap DET) = 0$, $p_{OD}(INF \cap DET) = 1$. Thus, we get

$$\begin{aligned} p_S(INF \cap DET) &= p_{SI}p_I(INF \cap DET) = \mathbb{P}(S \rightarrow I)\mathbb{P}(I \rightarrow OD) \\ &= \frac{\beta}{\beta + \delta} \mathcal{L}_{G_{I \rightarrow OD}}(\gamma) = \frac{\beta}{\beta + \delta} \frac{(\beta + \delta)e^{-(\beta + \delta)B}(e^{(\beta + \delta - \gamma)B} - 1)}{(\beta + \delta - \gamma)(1 - e^{-(\beta + \delta)B})}, \end{aligned}$$

so that

$$p_S(INF \cap \overline{DET}) = \frac{\beta}{\beta + \delta} \left[1 - \frac{(\beta + \delta)e^{-(\beta + \delta)B}(e^{(\beta + \delta - \gamma)B} - 1)}{(\beta + \delta - \gamma)(1 - e^{-(\beta + \delta)B})} \right].$$

We consider now some parameter values inspired from ref.²⁴, that describes an outbreak of *Methicillin-resistant Staphylococcus Aureus (MRSA)* that took place in The London Hospital in 1982. In particular, we consider an outbreak that was 16 weeks long, where $\frac{6}{103}$ healthcare workers were infected. We translate these quantities into

$$\begin{aligned} \delta &= \frac{1}{16} \text{weeks}^{-1}, \\ \frac{\beta}{\beta + \delta} &= \frac{6}{103}. \end{aligned}$$

From ref.²⁵, we consider that the healthcare worker under surveillance can spontaneously recover at rate $\gamma = 0.0475 \text{ weeks}^{-1}$. In Fig. 12 we compare the theoretical prediction for the process described by Fig. 11 and numerical results from the simulation of this process. Figure 12a shows the impact of the treatment on the time to recovery if we assume that the HCW has been infected (i.e., for the initial state I), measured by the increase in the recovery rate after entering the off-duty state (and the subsequent treatment), $\tilde{\gamma}/\gamma$, and for different values of B and M corresponding to the screening policies in place. It can be seen that the parameter M (days between screening after the healthcare worker has been sent off-duty) plays a more crucial role than the screening policy

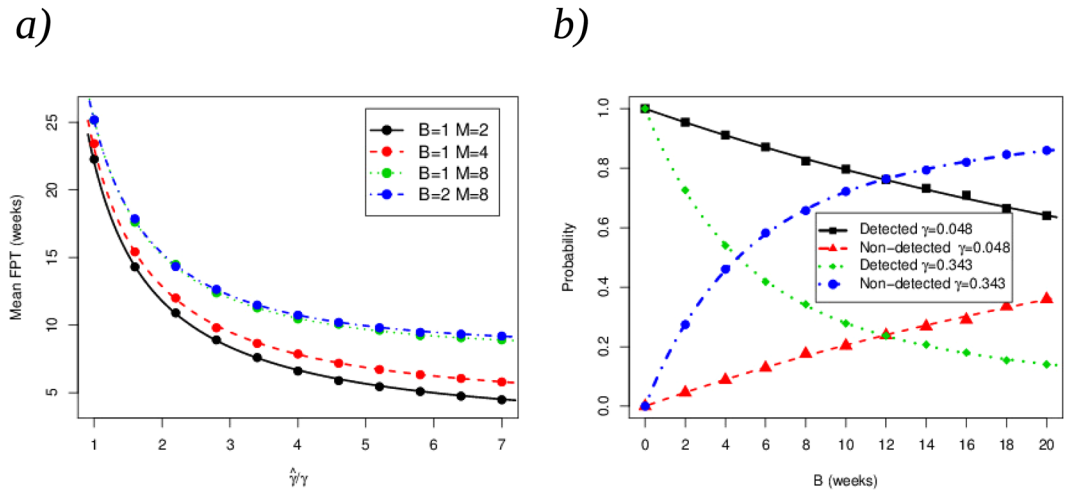


Figure 12. (a) Mean FPT to state Δ from the initial infected state I , for different screening policies (iv) and (v) described in terms of the parameters B and M . (b) Probabilities of detection and non-detection for the initial state I .

in place after the outbreak is detected (related to parameter B), when studying the mean FPT to state Δ . In all cases, the symbols represent the numerical simulations and the solid lines the theoretical prediction, respectively. In Fig. 12b, we plot the probabilities of detecting and not detecting the infection for an initially infected HCW,

$$p_I(INF \cap DET) \text{ and } p_I(INF \cap \overline{DET}),$$

for two different values of γ and for different values of B . Figure 12b shows that if the recovery occurs fast ($\gamma = 0.343 \text{ weeks}^{-1}$), a screening policy corresponding to small values of B (and requiring high levels of surveillance, and thus high costs) needs to be in place in order to ensure the detection of the infection for the HCW. On the other hand, detection probabilities close to one can still be obtained for relatively large values of B (requiring low levels of surveillance, and thus low costs) if the infectious period is large enough ($\gamma = 0.048 \text{ weeks}^{-1}$). For example, if $\gamma = 0.343 \text{ weeks}^{-1}$, the probability of the infection being undetected is near 50% for a policy corresponding to screening HCWs every $B = 4 \text{ weeks}$. On the other hand, this screening policy allows for detecting the infection with probability above 90% if $\gamma = 0.048 \text{ weeks}^{-1}$. This example illustrates the importance of setting screening policies according to both the infectivity and the infectivity period of the specific pathogen.

Discussion and Conclusion

In this paper, we make use of semi-Markov processes, where the absolute waiting time distribution from a given state to another is a function of the time since the system entered the first state. We have shown the relationship between the absolute waiting times to conditioned waiting times and the transition probabilities of the process. By means of first step arguments, we obtain the matrix $\mathbf{A}(z)$ in Eq. (1), which contains all the relevant information of the system and thus, allows for the computation of first passage times as well as the probabilities of different outcomes.

Exact computation can be carried out when all but one of the absolute waiting times for transitions (*i.e.*, the arrows) out of any given state follow an exponential distribution. When more than one absolute waiting time is non-exponential, we replace all but one of the non-exponential distributions by phase-type ones. Our approach has been shown to have applications in different biological contexts. For instance, as we have illustrated in case study 1 in the Results section, we can combine different experiments to obtain molecular rates (as, for instance, the generally distributed absolute waiting time for the phosphorylation events) or, as in case study 3, the impact of different screening strategies on the detection or end of a nosocomial outbreak.

Our approach relies on an analytical form of the Laplace-Stieltjes transform for each of the general distributions involved in the absolute waiting times. Hence, it cannot be used when the Laplace-Stieltjes transform is not well defined (e.g., the Cauchy distribution). However, of special relevance is the use of discrete absolute waiting time distributions for which the Laplace-Stieltjes transforms are well defined, as in case study 3, or deterministic (*i.e.*, degenerate) distributions (with $\mathcal{L}(z) = e^{-Tz}$ and T a fixed but fiducial time). This might be useful in laboratory experiments where the scientist regularly interacts with the experiment (for instance, adding ligand in a binding experiment) at pre-established discrete deterministic times.

From a computational perspective, for biological systems, the matrix $\mathbf{A}(z)$ is usually sparse, so it is amenable for inversion algorithms that are cost-effective and thus, allows one to extend the analysis presented in our case studies to larger systems. We note that for the case studies considered in this paper, no inversion algorithms were required. For a given system represented in terms of a discrete space with N states, where for example two events (*i.e.*, arrows) leaving each state are generally distributed, one can follow our approach and replace one of these generally distributed absolute waiting times, per state, by an M -dimensional phase-type distribution, which would lead to the analysis of an alternative stochastic process with MN states; similarly to our example in Fig. 9,

where $M = 3$. Although the resulting $A(z)$ might be sparse, so that efficient computational techniques could be in principle applied, computational limitations might still arise for significantly large values of N , when approximating more than one general distribution by a phase-type one for many events, or when high-dimensional phase-type approximations were needed (instead of the three-dimensional one used in Fig. 9, which might directly depend on the particular general distributions to be approximated). In the case of having an infinite number of states, the applicability of our approach relies on finding appropriate recursive solutions for the (infinite) system given by Eq. (1).

It is also worth mentioning that one natural generalisation of our approach is to study other quantities and not only first passage times, such as the probabilities of not becoming infected, becoming infected and being detected, or becoming infected but not being detected in case study 3), or in general any summary statistic or stochastic descriptor of the system allowing for a first step analysis approach. We note that the differences observed when studying first passage times for processes with generally distributed inter-event times, compared to their Markovian counterparts, could be even more significant when analysing different observables (stochastic descriptors). Analysing other stochastic descriptors is out of the scope of the present paper and will be the aim of our future collaborative effort.

As a final caveat, it is worth placing our approach in the context of the standard framework provided by ODE models, that are accurate as long as the probability of extinction is small, the population is large and the waiting time distributions of the underlying stochastic process are exponential. Without this latter assumption, the most accurate description is not an ordinary but an integro-differential equation (where the *kernel* of the distribution couples the value of the variables at the current time with their values in the past). Some authors have included this assumption (for instance, to model the eclipse phase in a viral infection) making use of delay differential equations²⁶. As shown in Fig. 10b, replacing only one of the events by a general distribution is not accurate unless the other competing times also accommodate non-exponential distributions^{27,28}.

Data Statement

Computer codes to generate the results derived in this paper are available online at Castro M, López-García M, Lythe G, Molina-París C. First passage events in biological systems with non-exponential inter-event times (computer codes). University of Leeds. [Dataset] <https://doi.org/10.5518/359>.

References

- Mckeithan, T. W. Kinetic proofreading in T-cell receptor signal transduction. *Proceedings of the national academy of sciences* **92**, 5042–5046 (1995).
- Wilkinson, D. J. *Stochastic modelling for systems biology*. (CRC press, 2011).
- Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry* **81**, 2340–2361 (1977).
- Gladnikoff, M. & Itay, R. Directly monitoring individual retrovirus budding events using atomic force microscopy. *Biophysical journal* **94**, 320–326 (2008).
- Yates, C. A., Ford, M. J. & Mort, R. L. A multi-stage representation of cell proliferation as a Markov process. *Bulletin of mathematical biology* **79**, 2905–2928 (2017).
- Duffy, K. R. & Hodgkin, P. D. Intracellular competition for fates in the immune system. *Trends in cell biology* **22**, 457–464 (2012).
- Limnios, N. & Oprisan, G. *Semi-Markov processes and reliability*. (Springer Science & Business Media, 2012).
- Huzurbazar, A. V. *Flowgraph models for multistate time-to-event data*, vol. 439 (John Wiley & Sons, 2004).
- Distefano, S. & Trivedi, K. S. Non-markovian state-space models in dependability evaluation. *Quality and Reliability Engineering International* **29**, 225–239 (2013).
- Janssen, J. *Semi-Markov models: theory and applications*. (Springer Science & Business Media, 2013).
- Iyer-Biswas, Z. *First passage processes in cellular biology*. Advances in Chemical Physics, Volume 160, First Edition. Stuart A. Rice and Aaron R. Dinner. (2016).
- Chow, T. & D'Orsogna, M. R. 'First passage problems in Biology', in *First passage phenomena and their applications*. Eds Metzler, R., Oshanin, G. & Redner, S. World Scientific 306–345 (2014).
- He, Q.-M. *Fundamentals of matrix-analytic methods*. (Springer, 2014).
- Zhang, H. The first passage times of phase semi-Markov processes. *Statistics & Probability Letters* **82**, 40–48 (2012).
- Abate, J. & Whitt, W. Computing Laplace transforms for numerical inversion via continued fractions. *INFORMS Journal on Computing* **11**, 394–405 (1999).
- Aslett, L. J. & Wilson, S. P. Markov chain monte carlo for inference on phase-type models. *Int. Statistical Inst.: Proc. 58th World Statistical Congress, 2011, Dublin (Session CPS027)* (2011).
- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/. Aslett, L. R package 'PhaseType', <https://cran.r-project.org/web/packages/PhaseType/index.html> (2013).
- Castro, M. *et al.* Receptor pre-clustering and T cell responses: Insights into molecular mechanisms. *Frontiers in Immunology* **5**, 132 (2014).
- Karlin, S. & Tavaré, S. Linear birth and death processes with killing. *Journal of Applied Probability* **19**, 477–487 (1982).
- Bel, G., Munsy, B. & Nemenman, I. The simplicity of completion time distributions for common complex biochemical processes. *Physical biology* **7**, 016003 (2009).
- Chylek, L. A. *et al.* Phosphorylation site dynamics of early t-cell receptor signaling. *PLoS One* **9**, e104240 (2014).
- Holcman, D. & Schuss, Z. The narrow escape problem. *SIAM Review* **56**, 213–257 (2014).
- Allen, L. J. *An introduction to stochastic processes with applications to biology*. (CRC Press, 2010).
- Duckworth, G., Lothian, J. & Williams, J. Methicillin-resistant staphylococcus aureus: report of an outbreak in a london teaching hospital. *Journal of Hospital Infection* **11**, 1–15 (1988).
- Wang, X., Panchanathan, S. & Chowell, G. A data-driven mathematical model of CA-MRSA transmission among age groups: evaluating the effect of control interventions. *PLoS Comput Biol* **9**, e1003328 (2013).
- Nelson, P. W. & Perelson, A. S. Mathematical analysis of delay differential equation models of HIV-1 infection. *Mathematical biosciences* **179**, 73–94 (2002).
- Castillo, E. *Functional equations and modelling in science and engineering*, vol. 161 (CRC Press, 1992).
- Horvath, A., Scarpa, M. & Telek, M. Phase type and matrix exponential distributions in stochastic modeling. In *Principles of Performance and Reliability Modeling and Evaluation*, 3–25 (Springer, 2016).

Acknowledgements

This work has been partially supported by grants FIS2013-47949-C2-2-P and FIS2016-78883-C2-2-P and PRX 16/00287 (MC), PIRSES-GA-2012-317893 (M.C., M.L.G., G.L. and C.M.P.), the Medical Research Council (UK) through a Skills Development Fellowship MR/N014855/1 (M.L.G.) and the Spanish Ministry of Economy and Competitiveness MTM2014-58091-P (M.C.). We thank A. Gómez-Corral, M. Telek and I. Adan for their valuable comments to previous versions of this manuscript.

Author Contributions

M.C. and M.L.G. conceived the idea. All authors contributed to developing the mathematical models, reviewing the literature, writing, reading and revising this manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-32961-7>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018