



Generalizations of the ‘Linear Chain Trick’: incorporating more flexible dwell time distributions into mean field ODE models

Paul J. Hurtado¹ · Adam S. Kiro Singh²

Received: 22 August 2018 / Revised: 10 June 2019 / Published online: 13 August 2019
© The Author(s) 2019

Abstract

In this paper we generalize the Linear Chain Trick (LCT; aka the Gamma Chain Trick) to help provide modelers more flexibility to incorporate appropriate dwell time assumptions into mean field ODEs, and help clarify connections between individual-level stochastic model assumptions and the structure of corresponding mean field ODEs. The LCT is a technique used to construct mean field ODE models from continuous-time stochastic state transition models where the time an individual spends in a given state (i.e., the dwell time) is Erlang distributed (i.e., gamma distributed with integer shape parameter). Despite the LCT’s widespread use, we lack general theory to facilitate the easy application of this technique, especially for complex models. Modelers must therefore choose between constructing ODE models using heuristics with oversimplified dwell time assumptions, using time consuming derivations from first principles, or to instead use non-ODE models (like integro-differential or delay differential equations) which can be cumbersome to derive and analyze. Here, we provide analytical results that enable modelers to more efficiently construct ODE models using the LCT or related extensions. Specifically, we provide (1) novel LCT extensions for various scenarios found in applications, including conditional dwell time distributions; (2) formulations of these LCT extensions that bypass the need to derive ODEs from integral equations; and (3) a novel Generalized Linear Chain Trick (GLCT) framework that extends the LCT to a much broader set of possible dwell time distribution assumptions, including the flexible *phase-type* distributions which can approximate distributions on \mathbb{R}^+ and can be fit to data.

Keywords Gamma chain trick · Linear Chain Trick · Distributed delay · Mean field model · Phase-type distributions · Time lag

✉ Paul J. Hurtado
phurtado@unr.edu

¹ University of Nevada, Reno, Reno, USA

² Stanford University, Stanford, USA

Mathematics Subject Classification 37N25 · 92B05 · 92D25 · 92D30 · 92D40 · 45D05 · 60J28

1 Introduction

Many scientific applications involve systems that can be framed as continuous time state transition models (e.g., see Strogatz 2014), and these are often modeled using mean field ordinary differential equations (ODE) of the form

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \boldsymbol{\theta}, t), \quad (1)$$

where $\mathbf{x}(t) \in \mathbb{R}^n$, parameters $\boldsymbol{\theta} \in \mathbb{R}^p$, and $f : \mathbb{R}^n \mapsto \mathbb{R}^n$ is smooth. The abundance of such applications, and the accessibility of ODE models in terms of the analytical techniques and computational tools for building, simulating, and analyzing ODE models [e.g., straightforward simulation methods and bifurcation analysis software like MatCont (Dhooge et al. 2003) and XPPAUT (Ermentrout 1987, 2019)], have made ODEs one of the most popular modeling frameworks in scientific applications.

Despite their widespread use, one shortcoming of ODE models is their inflexibility when it comes to specifying probability distributions that describe the duration of time individuals spend in a given state. The basic options available for assuming specific dwell time distributions within an ODE framework can really be considered as a single option: the 1st event time distribution for a (nonhomogeneous) Poisson process, which includes the exponential distribution as a special case.

To illustrate this, consider the following SIR model of infectious disease transmission by Kermack and McKendrick (1927),

$$\frac{d}{dt} S(t) = -\lambda(t) S(t) \quad (2a)$$

$$\frac{d}{dt} I(t) = \lambda(t) S(t) - \gamma I(t) \quad (2b)$$

$$\frac{d}{dt} R(t) = \gamma I(t) \quad (2c)$$

where $S(t)$, $I(t)$, and $R(t)$ correspond to the number of susceptible, infected, and recovered individuals in a closed population at time t and $\lambda(t) \equiv \beta I(t)$ is the per-capita infection rate (also called the *force of infection* by Anderson and May (1992) and others). This model can be thought of as the mean field model for some underlying stochastic state transition model where a large but finite number of individuals transition from state S to I to R [see the Appendix, Kermack and McKendrick (1927) for a derivation, and see Armbruster and Beck (2017), Banks et al. (2013), and references therein for examples of the convergence of stochastic models to mean field ODEs].

Although multiple stochastic models can yield the same mean field deterministic model, it is common to consider a stochastic model based on Poisson processes. For the SIR model above, for example, this stochastic analog would assume that, over the time

interval $[t, t + \Delta t]$ (for very small Δt), an individual in state S or I at time t is assumed to transition from S to I with probability $\lambda(t) \Delta t$, or from I to R with probability $\gamma \Delta t$, respectively. Taking $\Delta t \rightarrow 0$ yields the desired continuous time stochastic model. Here, the linear rate of transitions from I to R ($\gamma I(t)$) arises from assuming the dwell time for an individual in the infected state (I) follows an exponential distribution with rate γ (i.e., the 1st event time distribution for a homogeneous Poisson process with rate γ). Similarly, assuming the time spent in state S follows the 1st event time distribution under a nonhomogeneous (also called inhomogeneous) Poisson process with rate $\lambda(t)$ yields a time-varying per capita transition rate $\lambda(t)$. This association of a mean field ODE with a specific underlying stochastic model provides very valuable intuition in an applied context. For example, it allows modelers to ascribe application-specific (e.g., biological) interpretations to parameters and thus estimate parameter values (e.g., for γ above, the mean time spent infectious is $1/\gamma$), and it provides intuition and a clear mathematical foundation from which to construct and evaluate mean field ODE models based on individual-level, stochastic assumptions.

To construct similar models using other dwell time distributions, a standard approach is to formulate a continuous time stochastic model and from it derive mean field *distributed delay equations*, typically represented as integro-differential equations (IDEs) or integral equations (IEs) (e.g., see Kermack and McKendrick 1927; Hethcote and Tudor 1980; Feng et al. 2007, 2016, or see the example derivation in the Appendix). Readers unfamiliar with IEs and IDEs are referred to Burton (2005) or similar texts. IEs and IDEs have proven to be quite useful models in biology, e.g., they have been used to model chemical kinetics (Roussel 1996), gene expression (Smolen et al. 2000; Takashima et al. 2011; Guan and Ling 2018), physiological processes such as glucose-insulin regulation (Makroglou et al. 2006, and references therein), cell proliferation and differentiation (Özbay et al. 2008; Clapp and Levy 2015; Yates et al. 2017), cancer biology and treatment (Piotrowska and Bodnar 2018; Krzyzanski et al. 2018; Câmara De Souza et al. 2018), pathogen and immune response dynamics (Fenton et al. 2006), infectious disease transmission (Anderson and Watson 1980; Lloyd 2001a, b; Feng and Thieme 2000; Wearing et al. 2005; Lloyd 2009; Feng et al. 2007; Ciaravino et al. 2018), and population dynamics (MacDonald 1978a; Blythe et al. 1984; Metz and Diekmann 1986; Boese 1989; Nisbet et al. 1989; Cushing 1994; Wolkowicz et al. 1997; Gyllenberg 2007; Wang and Han 2016; Lin et al. 2018; Robertson et al. 2018). See also Campbell and Jessop (2009) and the applications reviewed therein.

However, while distributed delay equations are very flexible, in that they can incorporate arbitrary dwell time distributions, when compared to ODEs they also can be more challenging to derive, to analyze mathematically, and to simulate (Cushing 1994; Burton 2005). Thus, many modelers face a trade-off between building appropriate dwell time assumptions into their mean field models (i.e., opting for an IE or IDE model) and constructing parsimonious models that are more easily analyzed both mathematically and computationally (i.e., opting for an ODE model). For example, the following system of integral equations generalizes the SIR example above by incorporating an arbitrary distribution for the duration of infectiousness (i.e., the dwell time in state I):

$$S(t) = S(0)(1 - F_S(t)) \quad (3a)$$

$$I(t) = I(0)(1 - F_I(t)) + \int_0^t \beta I(u) S(u) (1 - F_I(u)) du \quad (3b)$$

$$R(t) = N - S(t) - I(t) \quad (3c)$$

where $N = S(0) + I(0) + R(0)$, $1 - F_S(t) = \exp(-\int_0^t \beta I(u) du)$ is the survival function for the distribution of time spent in susceptible state S [i.e., the 1st event time under a Poisson process with rate $\lambda(t) = \beta I(t)$], and $1 - F_I(t) = \exp(-\gamma t)$ is the survival function for the time spent in the infected state I (related models can be found in, e.g., Feng and Thieme 2000; Ma and Earn 2006; Krylova and Earn 2013; Champredon et al. 2018). A different choice of the CDF F_I allows us to generalize the SIR model to other dwell time distributions that describe the time individuals spend in the infected state. Integral equations like those above can also be differentiated (assuming the integrands are differentiable) and represented as integrodifferential equations (e.g., as in Hethcote and Tudor 1980).

There have been some efforts in the past to identify which categories of integral and integro-differential equations can be reduced to systems of ODEs (e.g., MacDonald 1989; Metz and Diekmann 1991; Ponosov et al. 2002; Jacquez and Simon 2002; Burton 2005; Goltser and Domoshnitsky 2013; Diekmann et al. 2017, and references therein), but in practice the most well known case is the reduction of IEs and IDEs that assume Erlang¹ distributed dwell times. This is done using what has become known as the Linear Chain Trick (LCT, also referred to as the Gamma Chain Trick; MacDonald 1978b; Smith 2010) which dates at least back to Fargue (1973) and earlier work by Theodore Vogel (e.g., Vogel 1961, 1965, according to Câmara De Souza et al. 2018). However, for more complex models that exceed the level of complexity that can be handled by existing “rules of thumb” like the LCT, the current approach is to derive mean field ODEs from mean field integral equations that might themselves first need to be derived from system-specific stochastic state transition models (e.g., Kermack and McKendrick 1927; Feng et al. 2007; Banks et al. 2013; Feng et al. 2016, and see the Appendix for an example.). Unfortunately, modelers often avoid these extra (often laborious) steps in practice by assuming (sometimes only implicitly) very simplistic dwell time distributions based on Poisson process 1st event times as in the SIR example above.

In light of the widespread use of ODE models, these challenges and trade-offs underscore a need for a more rigorous theoretical foundation to more effectively and more efficiently construct mean field ODE models that include more flexible dwell time distribution assumptions (Wearing et al. 2005; Feng et al. 2016; Robertson et al. 2018). The goal of this paper is to address these needs by (1) providing a theoretical foundation for constructing the desired system of ODEs directly from “first principles” (i.e., stochastic model assumptions), without the need to derive ODEs from intermediate IDEs or explicit stochastic models, and by (2) providing similar analytical results for novel extensions of the LCT which allow more flexible dwell time distributions, and conditional relationships among dwell time distributions, to be incorporated into

¹ Erlang distributions are Gamma distributions with integer-valued shape parameters.

ODE models. We also aim to clarify how underlying (often implicit) stochastic model assumptions are reflected in the structure of corresponding mean field ODE model equations.

The remainder of this paper is organized as follows. An intuitive description of the Linear Chain Trick (LCT) is given in Sect. 1.1 as a foundation for the extensions that follow. In Sect. 2 we review key notation and properties of Poisson processes and certain probability distributions needed for the results that follow. In Sect. 3.1.1 we highlight the association between Poisson process intensity functions and per capita rates in mean field ODEs, and in Sect. 3.1.2 we introduce what we call the *weak memorylessness* property of (nonhomogeneous) Poisson process 1st event time distributions. In Sects. 3.2 and 3.3 we give a formal statement of the LCT and in Sect. 3.4 a generalization that allows time-varying rates in the underlying Poisson processes. We then provide similar LCT generalizations for more complex cases: In Sect. 3.5 we provide results for multiple ways to implement transitions from one state to multiple states (which arise from different stochastic model assumptions and lead to different systems of mean field ODEs), and we address dwell times that obey Erlang mixture distributions. In Sect. 3.6 we detail how to construct mean field ODEs in which a sub-state transition (e.g., from infected to quarantined) does or does not alter the overall dwell time distribution in that state (e.g., the duration of infection). Lastly, in Sect. 3.7 we present a Generalized Linear Chain Trick (GLCT) which describes how to construct mean field ODEs from first principles based on assuming a very flexible family of dwell time distributions that include the *phase-type* distributions, i.e., hitting time distributions for continuous time Markov chains (Reinecke et al. 2012a; Horváth et al. 2016) which we address more specifically in Sect. 3.7.1. Tools for fitting phase-type distributions to data, or using them to approximate other distributions, are also mentioned in Sect. 3.7.1.

1.1 Intuitive description of the Linear Chain Trick

To begin, an intuitive understanding of the Linear Chain Trick (LCT) based on properties of Poisson processes, is helpful for drawing connections between underlying stochastic model assumptions and the structure of corresponding mean field ODEs. Here we consider a very basic case (detailed in Sect. 3.2): the mean field ODE model for a stochastic process in which particles in state X remain there for an Erlang(r, k) distributed amount of time before exiting to some other state.

In short, the LCT exploits a natural stage structure within state X imposed by assuming an Erlang(r, k) distributed dwell time [i.e., a gamma(r, k) distribution with integer shape k]. Recall that the time to the k th event under a homogeneous Poisson process with rate $r > 0$ is Erlang(r, k) distributed. In that context, each event is preceded by a length of time that is exponentially distributed with rate r , and thus the time to the k th event is the sum of k independent and identically distributed exponential random variables [i.e., the sum of k *iid* exponential random variables with rate r is Erlang(r, k) distributed]. Particles in state X at a given time can therefore be classified by which event they are awaiting next, i.e., a particle is in state X_i if it is waiting for the i th event to occur, and X_1, \dots, X_k is a partition of X . The dwell time distribution

for each sub-state X_i is exponential with rate r , and particles leave the last state X_k (and thus X) upon the occurrence of the k th event.

This sub-state partition is useful to impose on X because the corresponding mean field equations for these sub-states are linear (or nearly linear) ODEs. Specifically, if $x_i(t)$ is the amount in X_i at time t , these mean field ODEs are

$$\begin{aligned} \frac{d}{dt}x_1(t) &= -r x_1(t), \\ \frac{d}{dt}x_i(t) &= r x_{i-1}(t) - r x_i(t) \quad \text{for } i = 2, \dots, k \end{aligned} \quad (4)$$

where the total amount in X at time t is $x(t) = \sum_{i=1}^k x_i(t)$, $x_1(0) = x_0$ and $x_i(0) = 0$ for $i = 2, \dots, k$.

As we show below, this Poisson process based perspective allows us to generalize the LCT to other more complex scenarios where we partition a focal state X in a similar fashion. Those results are further extended from exponential dwell times to 1st event time distributions under a nonhomogeneous Poisson processes with time varying rate $r(t)$, allowing for time-varying (or state-dependent) dwell time distributions to be used in extensions of LCT.

2 Model framework

The context in which we consider applications of the Linear Chain Trick (LCT) is the derivation of continuous time mean field model equations for stochastic state transition models with a distributed dwell time in a focal state, X . Such mean field models might otherwise be modeled as integral equations (IEs) or integro-differential equations (IDEs), and we seek to identify generalizations of the LCT that allow us to replace mean field integral equations with equivalent systems of 1st order ODEs.

2.1 Distributions and notation

Below we will extend the LCT from Erlang(r, k) distributions (i.e., k th event time distributions under homogeneous Poisson processes with rate r) to event time distributions under nonhomogeneous Poisson processes with time varying rate $r(t)$, and related distributions like the minimum of multiple Erlang random variables. In this section we will first review some basic properties of these distributions.

Gamma distributions can be parameterized² by two strictly positive quantities: *rate* r and *shape* k (sometimes denoted α and β , respectively). The Erlang family of distributions are the subset of gamma distributions with integer-valued shape parameters $k \in \mathbb{Z}_+$, or equivalently, the distributions resulting from the sum of k iid exponential distributions. That is, if a random variable $T = \sum_{i=1}^k T_i$, where all T_i are independent exponential distributions with rate r , then T is Erlang(r, k) distributed. Since the

² They can also be parameterized in terms of their mean and variance (see Appendix B), or with a *shape* and *scale* parameters, where the *scale* parameter is the inverse of the *rate*.

inter-event times under a homogeneous Poisson process are exponentially distributed, the time to the k th event is thus Erlang(r, k). This construction is foundational to a proper intuitive understanding of the LCT and its extensions.

If random variable T is gamma(r, k) distributed, then its mean μ , variance σ^2 , and coefficient of variation c_v are given by

$$\mu = \frac{k}{r}, \sigma^2 = \frac{k}{r^2}, \text{ and } c_v = \frac{1}{\sqrt{k}}. \tag{5}$$

Note that gamma distributions can be parameterized by their mean μ and variance σ^2 by rewriting rate r and shape k using Eq. (5):

$$r = \frac{\mu}{\sigma^2}, \text{ and } k = \frac{\mu^2}{\sigma^2}. \tag{6}$$

However, to ensure this gamma distribution is also Erlang (i.e., to ensure the shape parameter k is an integer) one must adjust the assumed variance up or down by rounding the value of k in Eq. (6) down or up, respectively, to the nearest integer (see Appendix B for details, and alternatives).

The Erlang density function (g), CDF (G), and survival³ function ($S = 1 - G$; also called the *complementary CDF*) are given by

$$g_r^k(t) = r \frac{(rt)^{k-1}}{(k-1)!} e^{-rt} \tag{7a}$$

$$G_r^k(t) = 1 - \sum_{j=1}^k \frac{(rt)^{j-1}}{(j-1)!} e^{-rt} = 1 - \sum_{j=1}^k \frac{1}{r} g_r^j(t) \tag{7b}$$

$$S_r^k(t) = 1 - G_r^k(t) = \sum_{j=1}^k \frac{1}{r} g_r^j(t). \tag{7c}$$

The results in Sect. 3 use (and generalize) the property of Erlang distributions detailed in Lemma 1 (eqs. 7.11 in Smith 2010, restated here without proof), which is the linchpin of the LCT.

Lemma 1 *Erlang density functions $g_r^j(t)$, with rate r and shape j , satisfy*

$$\frac{d}{dt} g_r^1(t) = -r g_r^1(t), \quad \text{where } g_r^1(0) = r, \tag{8a}$$

$$\frac{d}{dt} g_r^j(t) = r [g_r^{j-1}(t) - g_r^j(t)], \quad \text{where } g_r^j(0) = 0 \text{ for } j \geq 2. \tag{8b}$$

Since homogeneous Poisson processes are a special case of nonhomogeneous Poisson processes⁴ from here on we will use ‘‘Poisson process’’ or ‘‘Poisson process with

³ A useful interpretation of survival functions, which is used below, is that they give the expected proportion remaining after a give amount time.

⁴ ...despite the implied exclusivity of the adjective *nonhomogeneous*.

rate $r(t)$ ” to refer to cases that apply to both homogeneous [i.e., $r(t) = r$ constant] and nonhomogeneous Poisson processes.

The more general k th event time distribution under a Poisson process with rate $r(t)$, starting from some time $\tau < t$ has a density function (h_r^k), survival function (S_r^k), and CDF ($H_r^k \equiv 1 - S_r^k$) given by

$$h_r^k(t, \tau) = r(t) \frac{m(t, \tau)^{k-1}}{(k-1)!} e^{-m(t, \tau)} \quad \text{and} \quad (9a)$$

$$S_r^k(t, \tau) = \sum_{j=1}^k \frac{h_r^j(t, \tau)}{r(t)} \quad \text{where} \quad (9b)$$

$$m(t, \tau) \equiv \int_{\tau}^t r(s) ds \quad \text{and} \quad \frac{d}{dt} m(t, \tau) = r(t). \quad (9c)$$

For an arbitrary survival function starting at time τ (i.e., over the period $[\tau, t]$ where $t \geq \tau$) we use the notation $S(t, \tau)$. In some instances, we also use the notation $S(t) \equiv S(t, 0)$.

Lastly, in the context of state transitions models, it is common to assume that, upon leaving a given state (e.g., state X) at time t , individuals are distributed across multiple recipient states according to a generalized Bernoulli distribution (also known as the *categorical distribution* or the multinomial distribution with 1 trials) defined on the integers 1 through k where the probability of a particle entering the j th of k recipient states ($j \in 1, \dots, k$) is $p_j(t)$, which define probability vector $\mathbf{p}(t) = [p_1(t), \dots, p_k(t)]^T$ and $\sum_{j=1}^k p_j(t) = 1$.

3 Results

The results below focus on one or more states, within a potentially larger state transition model, for which we would like to assume a particular dwell time distribution and derive a corresponding system of mean field ODEs using the LCT or a generalization of the LCT. In particular, the results below describe how to construct those mean field ODEs directly from stochastic model assumptions without needing to derive them from equivalent mean field integral equations (which themselves might need to be derived from an explicit continuous-time stochastic model).

3.1 Preliminaries

Before presenting extensions of the LCT, we first illustrate in Sect. 3.1.1 how mean field ODEs include terms that reflect underlying Poisson process rates. In Sect. 3.1.2, we highlight a key property of these Poisson process 1st event time distributions that we refer to as a *weak memorylessness property* since it is a generalization of the well known memorylessness property of the exponential and geometric distributions.

3.1.1 Transition rates in ODEs reflect underlying Poisson process rates

To build upon the intuition spelled out in Sect. 1.1, where particles exit X following an exponentially distributed dwell time, we now assume that particles exit X following the 1st event time under nonhomogeneous Poisson processes with rate $r(t)$ (recall the 1st event time distribution is exponential if $r(t) = r$ is constant). As illustrated by the corresponding mean field equations given below, the rate function $r(t)$ can be viewed as either the intensity function⁵ for the Poisson process governing when individuals leave state X, or as the (mean field) per-capita rate of loss from state X as shown in Eq. (11). This dual perspective provides valuable intuition for critically evaluating mean field ODEs.

Example 1 (*Equivalence between Poisson process rates and per capita rates in mean field ODEs*) Consider the scenario described above. The survival function for the dwell time distribution for a particle entering X at time τ is $S(t, \tau) = \exp(-\int_{\tau}^t r(u) du)$, and it follows that the expected proportion of such particles remaining in X at time $t > \tau$ is given by $S(t, \tau)$. Let $x(t)$ be the total amount in state X at time t , $x(0) = x_0$, and assume that $\mathcal{I}(t)$ and $r(t)$ are integrable, non-negative functions of t . Then the corresponding mean field integral equation for this scenario is

$$x(t) = x_0 S(t, 0) + \int_0^t \mathcal{I}(\tau) S(t, \tau) d\tau \tag{10}$$

and the integral equation (10) above is equivalent to

$$\frac{d}{dt}x(t) = \mathcal{I}(t) - r(t)x(t), \text{ with } x(0) = x_0. \tag{11}$$

Proof The Leibniz rule for differentiating integrals, Eq. (10), and Lemma 1 yield

$$\begin{aligned} \frac{d}{dt}x(t) &= x_0 \frac{d}{dt}S(t, 0) + \frac{d}{dt} \int_0^t \mathcal{I}(\tau) S(t, \tau) d\tau \\ &= -r(t)x_0 e^{-\int_0^t r(u) du} + \mathcal{I}(t) - r(t) \int_0^t \mathcal{I}(\tau) e^{-\int_{\tau}^t r(u) du} d\tau \\ &= \mathcal{I}(t) - r(t) \left[x_0 e^{-\int_0^t r(u) du} + \int_0^t \mathcal{I}(\tau) e^{-\int_{\tau}^t r(u) du} d\tau \right] \\ &= \mathcal{I}(t) - r(t)x(t). \end{aligned} \tag{12}$$

□

3.1.2 Weak memoryless property of Poisson process 1st event time distributions

The intuition behind the LCT, and the history-independent nature of ODEs, are related to the memorylessness property of exponential distributions. For example, when par-

⁵ That is, the probability of a given individual exiting state X during a brief time period $[t, t + \Delta t]$ is approximately $r(t)\Delta t$.

ticles accumulate in a state with an exponentially distributed dwell time distribution, then at any given time all particles currently in that state have *iid* exponentially distributed amounts of time left before they leave that state regardless of the duration of time already spent in that state. Each of these can be viewed as a reflection of the history-independent nature of Poisson processes.

Accordingly, the familiar memorylessness property of exponential and geometric distributions can, in a sense, be generalized to (nonhomogeneous) Poisson process 1st event time distributions. Recall that if an exponentially distributed (rate r) random variable T represents the time until some event, then if the event has not occurred by time s the remaining duration of time until the event occurs is also exponential with rate r . The analogous *weak memorylessness* property of nonhomogeneous Poisson process 1st event time distributions is detailed in the following definition.

Definition 1 (Weak memorylessness property of Poisson process 1st event times)

Assume T is a Poisson process 1st event time starting at time τ , which has rate $r(t)$ and CDF $H_r^1(t, \tau) = 1 - \exp(-m(t, \tau))$ [see Eqs. (9) and (9c)]. If the event has not occurred by time $s > \tau$ the distribution of the remaining time $T_s \equiv T - s \mid T > s$ follows a shifted but otherwise identical Poisson process 1st event time distribution with CDF $P(T_s \leq t) = H_r^1(t + s, s)$. If $r(t) = r$ is a positive constant we recover the memorylessness property of the exponential distribution.

Proof The CDF of T_s (for $t > \tau$) is given by

$$\begin{aligned}
 P(T_s \leq t) &= P(T - s \leq t \mid T > s) = \frac{P(s < T \leq s + t)}{P(s < T)} \\
 &= \frac{H_r^1(t + s, \tau) - H_r^1(s, \tau)}{1 - H_r^1(s, \tau)} = 1 - \frac{1 - H_r^1(t + s, \tau)}{1 - H_r^1(s, \tau)} \tag{13} \\
 &= 1 - \frac{e^{-m(t+s,\tau)}}{e^{-m(s,\tau)}} = 1 - e^{-m(t+s,s)} = H_r^1(t + s, s).
 \end{aligned}$$

□

In other words, Poisson process 1st event time distributions are memoryless up to a time shift in their rate functions. In the context of multiple particles entering a given state X at different times and leaving according to independent Poisson process 1st event times with identical rates $r(t)$ (i.e., t is absolute time, not time since entry into X), then for all particles in state X at a given time the distribution of time remaining in state X is (1) independent of how much time each particle has already spent in X and (2) follows *iid* Poisson process 1st event time distributions with rate $r(t)$.

3.2 Simple case of the LCT

To illustrate how the LCT follows from Lemma 1, consider the simple case of the LCT illustrated in Fig. 1, where a higher dimensional model includes a state transition into, then out of, a focal state X. Assume the time spent in that state (T_X) is Erlang(r, k)

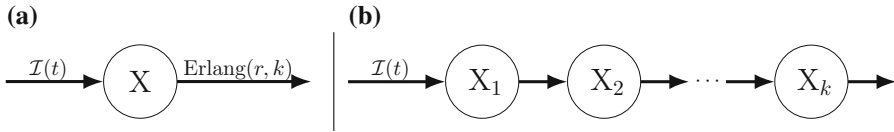


Fig. 1 Example diagram where state X has an Erlang(r, k) distributed dwell time, represented either as **a** a single state and corresponding integral equation, or **b** as a set of k sub-states each with exponential dwell time distributions whose mean field equations can be represented as either integral equations or a system of ODEs (see Theorem 1). Rate $\mathcal{I}(t)$ is an integrable non-negative function describing the mean field influx rate into state X

distributed [i.e., $T_X \sim \text{Erlang}(r, k)$]. Then the LCT provides a system of ODEs equivalent to the mean field integral equations for this process as detailed in the following theorem:

Theorem 1 (Simple LCT) *Consider a continuous time state transition model with inflow rate $\mathcal{I}(t) \geq 0$ (an integrable function of t) into state X which has an Erlang(r, k) distributed dwell time [with survival function S_r^k from Eq. (7c)]. Let $x(t)$ be the (mean field) amount in state X at time t and assume $x(0) = x_0$.*

The mean field integral equation for this scenario (see Fig. 1a) is

$$x(t) = x_0 S_r^k(t) + \int_0^t \mathcal{I}(s) S_r^k(t - s) ds. \tag{14}$$

State X can be partitioned (see Fig. 1b) into k sub-states $X_i, i = 1, \dots, k$, where particles in X_i are those awaiting the i th event as the next event under a homogeneous Poisson process with rate r . Let $x_i(t)$ be the amount in X_i at time t , and $x(t) = \sum_{j=1}^k x_j(t)$. Equation (14) is equivalent to the mean field ODEs

$$\frac{d}{dt} x_1(t) = \mathcal{I}(t) - r x_1(t) \tag{15a}$$

$$\frac{d}{dt} x_j(t) = r x_{j-1}(t) - r x_j(t), \quad j = 2, \dots, k \tag{15b}$$

with initial conditions $x_1(0) = x_0, x_j(0) = 0$ for $j \geq 2$, and

$$x_j(t) = x_0 \frac{1}{r} g_r^j(t) + \int_0^t \mathcal{I}(s) \frac{1}{r} g_r^j(t - s) ds. \tag{16}$$

Proof Substituting Eq. (7c) into Eq. (14) and then substituting Eq. (16) yields

$$\begin{aligned}
 x(t) &= x_0 S_r^k(t) + \int_0^t \mathcal{I}(s) S_r^k(t-s) ds \\
 &= x_0 \sum_{j=1}^k \frac{1}{r} g_r^j(t) + \int_0^t \mathcal{I}(s) \sum_{j=1}^k \frac{1}{r} g_r^j(t-s) ds \\
 &= \sum_{j=1}^k \left(x_0 \frac{1}{r} g_r^j(t) + \int_0^t \mathcal{I}(s) \frac{1}{r} g_r^j(t-s) ds \right) = \sum_{j=1}^k x_j(t).
 \end{aligned}
 \tag{17}$$

Differentiating Eq. (16) (for $j = 1, \dots, k$) yields Eq. (15) as follows.

For $j = 1$, Eq. (16) reduces to

$$x_1(t) = x_0 e^{-rt} + \int_0^t \mathcal{I}(s) e^{-r(t-s)} ds.
 \tag{18}$$

Differentiating $x_1(t)$ using the Leibniz integral rule and substituting (18) yields

$$\frac{d}{dt} x_1(t) = -r x_0 e^{-rt} - r \int_0^t \mathcal{I}(s) e^{-r(t-s)} ds + \mathcal{I}(t) = \mathcal{I}(t) - r x_1(t).
 \tag{19}$$

Similarly, for $j \geq 2$, Lemma 1 yields

$$\begin{aligned}
 \frac{d}{dt} x_j(t) &= x_0 \frac{1}{r} \frac{d}{dt} g_r^j(t) + \int_0^t \mathcal{I}(s) \frac{d}{dt} \left(\frac{1}{r} g_r^j(t-s) \right) ds \\
 &= x_0 \left(g_r^{j-1}(t) - g_r^j(t) \right) + \int_0^t \mathcal{I}(s) \left(g_r^{j-1}(t-s) - g_r^j(t-s) \right) ds \\
 &= r \left(\frac{x_0}{r} g_r^{j-1}(t) + \int_0^t \mathcal{I}(s) \frac{1}{r} g_r^{j-1}(t-s) ds \right) - r \left(\frac{x_0}{r} g_r^j(t) \right. \\
 &\quad \left. + \int_0^t \mathcal{I}(s) \frac{1}{r} g_r^j(t-s) ds \right) = r x_{j-1}(t) - r x_j(t).
 \end{aligned}
 \tag{20}$$

□

The X_j dwell time distributions are exponential with rate r . To see why, let $\chi_i(t)$ (where $1 \leq i \leq k$) be the (mean field) number of particles in state X at time t that have not reached the i th event. Then

$$\chi_i(t) = x_0 S_r^i(t) + \int_0^t \mathcal{I}(s) S_r^i(t-s) ds
 \tag{21}$$

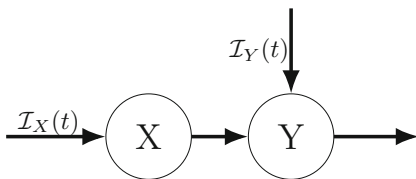
and by Eq. (7) we see from Eqs. (18) and (21) that $x_j(t) = \chi_j(t) - \chi_{j-1}(t)$. That is, particles in state X_j are those for which the $(j - 1)$ th event has occurred, but not the j th event. Thus, by properties of Poisson processes the dwell time in state X_j must exponential with rate r .

3.3 Standard LCT

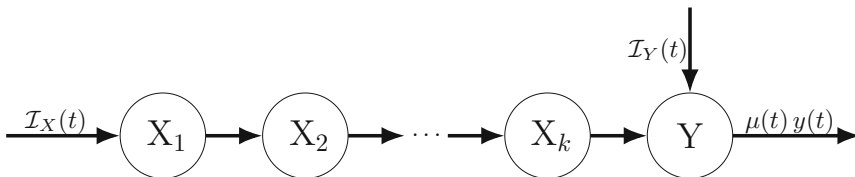
The following theorem and corollary together provide a more general, formal statement of the standard Linear Chain Trick (LCT) as used in practice. These extend Theorem 1 (compare Figs. 1, 2) to explicitly include that particles leaving X enter state Y then remain in Y according to an arbitrary distribution with survival function S , and include transitions into Y from other sources.

Theorem 2 (Standard LCT) *Consider a continuous time dynamical system model of mass transitioning among various states, with inflow rate $\mathcal{I}_X(t)$ to a state X and an Erlang(r, k) distributed delay before entering state Y. Let $x(t)$ and $y(t)$ be the amount in each state, respectively, at time t . Further assume an inflow rate $\mathcal{I}_Y(t)$ into state Y from other non-X states, and that the underlying stochastic model assumes that the duration of time spent in state Y is determined by survival function $S_Y(t, \tau)$. Assume $\mathcal{I}_i(t)$ are integrable non-negative functions of t , and assume non-negative initial conditions $x(0) = x_0$ and $y(0) = y_0$.*

(a)



(b)



(c)

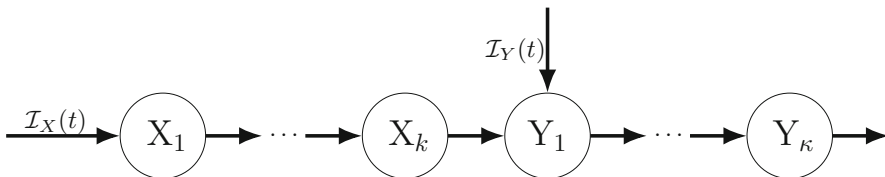


Fig. 2 (Standard LCT results) This generic case assumes that the dwell times in state X (see panel a) are Erlang(r, k) distributed with inflow rates $\mathcal{I}_X(t) \geq 0$ into state X and $\mathcal{I}_Y(t) \geq 0$ into state Y. Panels b and c show sub-states resulting from applying the LCT and Corollary 1 assuming either (b) dwell times in state Y are determined by per-capita rate function $\mu(t)$, or (c) dwell times in Y follow an Erlang distribution with shape parameter κ

The mean field integral equations for this scenario are

$$x(t) = x_0 S_r^k(t) + \int_0^t \mathcal{I}_X(s) S_r^k(t - s) ds \tag{22a}$$

$$y(t) = y_0 S_Y(t, 0) + \int_0^t \left(\mathcal{I}_Y(\tau) + x_0 g_r^k(\tau) + \int_0^\tau \mathcal{I}_X(s) g_r^k(\tau - s) ds \right) S_Y(t, \tau) d\tau. \tag{22b}$$

Equations (22) are equivalent to

$$\frac{d}{dt} x_1(t) = \mathcal{I}_X(t) - r x_1(t) \tag{23a}$$

$$\frac{d}{dt} x_j(t) = r x_{j-1}(t) - r x_j(t), \quad j = 2, \dots, k \tag{23b}$$

$$y(t) = y_0 S_Y(t, 0) + \int_0^t \underbrace{(\mathcal{I}_Y(\tau) + r x_k(\tau))}_{\text{Net input rate at time } \tau} S_Y(t, \tau) d\tau \tag{23c}$$

where $x(t) = \sum_{j=1}^k x_j(t)$, $x_1(0) = x_0$, $x_j(0) = 0$ for $j \geq 2$, and

$$x_j(t) = x_0 \frac{1}{r} g_r^j(t) + \int_0^t \mathcal{I}_X(s) \frac{1}{r} g_r^j(t - s) ds. \tag{24}$$

Proof Eqs. (23a), (23b) and (24) follow from Theorem 1. Equation (23c) follows from substituting (24) into (22b). The definition of x_j and initial condition $x(0) = x_0$ together imply $x_1(0) = x_0$ and $x_j(0) = 0$ for the remaining $j \geq 2$. \square

Corollary 1 Integral equations like Eq. (23c) may have ODE representations depending on the Y dwell time distribution, i.e., $S_Y(t, \tau)$, for example:

1. If particles leave Y after a Poisson process 1st event time distributed dwell time [i.e., the per-capita loss rate from Y is $\mu(t)$], then $S_Y(t, \tau) = \exp(-\int_\tau^t \mu(u) du)$, and letting $\mathcal{I}(t) = \mathcal{I}_Y(t) + r x_k(t)$, Theorem 2 yields

$$\frac{d}{dt} y(t) = \mathcal{I}_Y(t) + r x_k(t) - \mu(t) y(t). \tag{25}$$

2. If particles leave Y after an Erlang(μ, κ) delay, then $S_Y(t, \tau) = S_\mu^\kappa(t - \tau)$ and letting $\mathcal{I}(t) = \mathcal{I}_Y(t) + r x_k(t)$ Theorem 2 gives that $y = \sum_{i=1}^\kappa y_i$ and

$$\frac{d}{dt} y_1(t) = \mathcal{I}_Y(t) + r x_k(t) - \mu y_1(t) \tag{26a}$$

$$\frac{d}{dt} y_i(t) = \mu y_{i-1}(t) - \mu y_i(t), \quad i = 2, \dots, \kappa. \tag{26b}$$

3. As implied by 1 and 2 above, if the per-capita loss rate $\mu(t) = \mu$ is constant or time spent in Y is otherwise exponentially distributed, then

$$\frac{d}{dt}y(t) = \mathcal{I}_Y(t) + rx_k(t) - \mu y(t). \tag{27}$$

4. Any of the more general cases considered in the sections below.

Example 2 To illustrate how the Standard LCT can be used to substitute an implicit exponential dwell time distribution with an Erlang distribution, consider the SIR example discussed in the Introduction [Eqs. (2) and (3), see also Anderson and Watson 1980; Lloyd 2001a, b], but assume the dwell time distribution for the infected state I is Erlang (still with mean $1/\gamma$) with variance⁶ σ^2 , i.e., by Eq. (6), Erlang with a rate $r = \gamma k$ and shape $k = \sigma^2/\gamma^2$.

By Theorem 2 and Corollary 1, with $\mathcal{I}_I(t) = \lambda(t) S(t)$ and $I(t) = \sum_{j=1}^k I_j(t)$, the corresponding mean field ODEs are

$$\frac{d}{dt}S(t) = -\lambda(t) S(t) \tag{28a}$$

$$\frac{d}{dt}I_1(t) = \lambda(t) S(t) - \gamma k I_1(t) \tag{28b}$$

$$\frac{d}{dt}I_j(t) = \gamma k I_{j-1}(t) - \gamma k I_j(t), \quad \text{for } j = 2, \dots, k \tag{28c}$$

$$\frac{d}{dt}R(t) = \gamma k I_k(t). \tag{28d}$$

Notice that if $\sigma^2 = \gamma^2$ (i.e., $k = 1$), the dwell time in I is exponentially distributed with rate γ , $I(t) = I_1(t)$, and Eq. (28) reduce to Eq. (2).

This example nicely illustrates how using Theorem 2 to relax an exponential dwell time assumption implicit in a system of mean field ODEs is much more straightforward than constructing them after first deriving the integral equations, like Eq. (3), and then differentiating them using Lemma 1. In the sections below, we present similar theorems intended to be used for constructing mean field ODEs directly from stochastic model assumptions.

3.4 Extended LCT for Poisson process k th event time distributed dwell times

The Standard LCT assumes an Erlang(r, k) distributed dwell time, i.e., a k th event time distribution under a homogeneous Poisson process with rate r . Here we generalize the Standard LCT by assuming the X dwell time follows a more general k th event time distribution under a Poisson process with rate $r(t)$.

⁶ Here the variance is assumed to have been chosen so that the resulting shape parameter is integer valued. See Appendix B for related details.

First, observe that Eq. (8) in Lemma 1 are more practical when written in terms of $\frac{1}{r}g_r^j(t)$ (see the proof of Theorem 1),

$$\frac{d}{dt} \left[\frac{1}{r}g_r^1(t) \right] = -r \left[\frac{1}{r}g_r^1(t) \right], \tag{29a}$$

$$\frac{d}{dt} \left[\frac{1}{r}g_r^j(t) \right] = r \left[\frac{1}{r}g_r^{j-1}(t) - \frac{1}{r}g_r^j(t) \right], \quad j \geq 2 \tag{29b}$$

where $\frac{1}{r}g_r^1(0) = 1$ and $\frac{1}{r}g_r^j(0) = 0$.

Lemma 2 *A similar relationship to Eq. (29) above (i.e., to Lemma 1) holds true for the Poisson process j th event time distribution density functions h_r^j given by Eq. (9a). Specifically,*

$$\frac{d}{dt} \left[\frac{1}{r(t)}h_r^1(t, \tau) \right] = -r(t) \left[\frac{1}{r(t)}h_r^1(t, \tau) \right], \tag{30a}$$

$$\frac{d}{dt} \left[\frac{1}{r(t)}h_r^j(t, \tau) \right] = r(t) \left[\frac{1}{r(t)}h_r^{j-1}(t, \tau) - \frac{1}{r(t)}h_r^j(t, \tau) \right], \tag{30b}$$

where $\frac{1}{r(\tau)}h_r^1(\tau, \tau) = 1$ and $\frac{1}{r(\tau)}h_r^j(\tau, \tau) = 0$ for $j \geq 2$. Note that, if for some t $r(t) = 0$, this relationship can be written in terms of

$$u_r^k(t, \tau) \equiv \frac{m(t, \tau)^{k-1}}{(k-1)!} e^{-m(t, \tau)}, \tag{31}$$

as shown in the proof below, where $h_r^k(t, \tau) = r(t)u_r^k(t, \tau)$, $u_r^1(\tau, \tau) = 1$, and $u_r^j(\tau, \tau) = 0$ for $j \geq 2$.

Proof For $j = 1$,

$$\frac{d}{dt} \left[u_r^1(t, \tau) \right] = \frac{d}{dt} e^{-m(t, \tau)} = -r(t) e^{-m(t, \tau)} = -r(t) u_r^1(t, \tau). \tag{32}$$

Likewise, for $j \geq 2$, we have

$$\begin{aligned} \frac{d}{dt} \left[u_r^j(t, \tau) \right] &= \frac{d}{dt} \frac{m(t, \tau)^{k-1}}{(k-1)!} e^{-m(t, \tau)} \\ &= r(t) \frac{m(t, \tau)^{k-2}}{(k-2)!} e^{-m(t, \tau)} - r(t) \frac{m(t, \tau)^{k-1}}{(k-1)!} e^{-m(t, \tau)} \\ &= r(t) \left[u_r^{j-1}(t, \tau) - u_r^j(t, \tau) \right]. \end{aligned} \tag{33}$$

□

Lemma 2 allows us to generalize Erlang-based results like Theorem 2 to their time-varying counterparts with a time-dependent (or state-dependent) rate $r(t)$, as in the following generalization of the Standard LCT (Theorem 2).

Theorem 3 (Extended LCT for dwell times distributed as Poisson process k^{th} event times) *Consider the Standard LCT in Theorem 2 but assume the X dwell time is a Poisson process k^{th} event time, rate $r(t)$. The corresponding mean field integral equations, where h_r^j and S_r^j are given in Eq. (9), are*

$$x(t) = x_0 S_r^k(t, 0) + \int_0^t \mathcal{I}_X(s) S_r^k(t, s) ds \tag{34a}$$

$$y(t) = y_0 S_Y(t, 0) + \int_0^t \left(\mathcal{I}_Y(\tau) + x_0 h_r^k(\tau, 0) + \int_0^\tau \mathcal{I}_X(s) h_r^k(\tau, s) ds \right) S_Y(t, \tau) d\tau. \tag{34b}$$

The above Eq. (34), with $x(t) = \sum_{j=1}^k x_j(t)$, are equivalent to

$$\frac{d}{dt} x_1(t) = \mathcal{I}_X(t) - r(t) x_1(t) \tag{35a}$$

$$\frac{d}{dt} x_j(t) = r(t) x_{j-1}(t) - r(t) x_j(t), \quad j = 2, \dots, k \tag{35b}$$

$$y(t) = y_0 S_Y(t, 0) + \int_0^t (\mathcal{I}_Y(\tau) + r(\tau) x_k(\tau)) S_Y(t, \tau) d\tau \tag{35c}$$

with initial conditions $x_1(0) = x_0, x_j(0) = 0$ for $j \geq 2$ and

$$x_j(t) = x_0 \frac{1}{r(t)} h_r^j(t, 0) + \int_0^t \mathcal{I}_X(s) \frac{1}{r(t)} h_r^j(t, s) ds. \tag{36}$$

Equation (35c) may be further reduced to ODEs, e.g., via Corollary 1.

Proof Substituting Eq. (9b) into Eq. (34a) and substituting Eq. (36) yields $x(t) = \sum_{j=1}^k x_j(t)$. Differentiating Eq. (36) with $j = 1$ using the Leibniz integration rule as well as Eq. (30a) from Lemma 2 yields Eq. (35a). Likewise, for $j \geq 2$, differentiation of Eq. (36) and Lemma 2 yields

$$\begin{aligned} \frac{d}{dt} x_j(t) &= x(0) r(t) \left[\frac{1}{r(t)} h_r^{j-1}(t, 0) - \frac{1}{r(t)} h_r^j(t, 0) \right] \\ &\quad + \int_0^t \mathcal{I}_X(s) r(t) \left[\frac{1}{r(t)} h_r^{j-1}(t, \tau) - \frac{1}{r(t)} h_r^j(t, \tau) \right] ds \\ &= r(t) (x_{j-1}(t) - x_j(t)). \end{aligned} \tag{37}$$

Equation (35c) follows from substituting (36) into (34b). The definition of x_j and initial condition $x(0) = x_0$ together imply $x_1(0) = x_0$ and $x_j(0) = 0$ for the remaining

$j \geq 2$. If $r(t) = 0$ for some t , Eq. (35) still hold, since Eqs. (36) and (37) can be rewritten using u as in the proof of Lemma 2. \square

Having generalized the Standard LCT (Lemma 1 and Theorem 2) to include Poisson process k th event time distributed dwell times, we may now address more complex stochastic model assumptions and how they are reflected in the structure of corresponding mean field ODEs.

3.5 Transitions to multiple states

Modeling the transition from one state to multiple states following a distributed delay (as illustrated in Fig. 3) can be done under different sets of assumptions about the underlying stochastic processes, particularly with respect to the rules governing how individuals are distributed across multiple recipient states upon exiting X , and how those rules depend on the dwell time distribution(s) for individuals in that state. Importantly, those different sets of assumptions can yield very different mean field models (e.g., see Feng et al. 2016) and so care must be taken to make those assumptions appropriately for a given application.

While modelers have some flexibility to choose appropriate assumptions, in practice modelers sometimes make unintended and undesirable implicit assumptions, especially when constructing ODE models using “rules of thumb” instead of deriving them from first principles. In this section we present results aimed at helping guide (a) the process of picking appropriate dwell time distribution assumptions, and (b) directly constructing corresponding systems of ODEs without deriving them from explicit stochastic models or intermediate integral equations.

Each of the three cases detailed below yield different mean field ODE models for the scenario depicted in Fig. 3.

First, in Sect. 3.5.1, we consider the extension of Theorem 3 where upon leaving X particles are distributed across $m \geq 1$ recipient states according to a generalized Bernoulli distribution with (potentially time varying) probabilities/proportions $p_j(t)$, $j = 1, \dots, m$. Here the outcome of which state a particle transitions to is independent of the time spent in the first state.

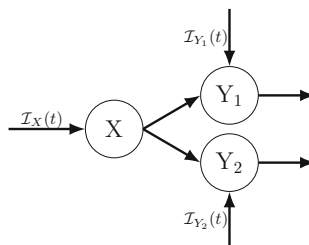


Fig. 3 Example diagram of transitions out of a given state (X) and into multiple states (Y_1 and Y_2). Different assumptions about (1) the dwell times in X , and (2) rules governing the subsequent transitions to Y_1 and/or Y_2 will lead to different sub-state partitions of X , and thus different mean field equations, as detailed in Sect. 3.5. Fortunately, different scenarios often encountered in applications can be reduced to ODEs by applying the results in Sect. 3.5 as detailed in Theorems 4, 5, and 6 and as illustrated in Figs. 4, 5, and 6

Second, in Sects. 3.5.2 and 3.5.3, particles entering the first state (X) do not all follow the same dwell time distribution in X. Instead, upon entering X they are distributed across $n \geq 2$ sub-states of X, X_i , according to a generalized Bernoulli distribution, and each sub-state X_i has a dwell time given by a Poisson process k_i th event time distribution with rate $r_i(t)$. That is, the X dwell time is a finite mixture of Poisson process event time distributions. Particles transition out of X into m subsequent states Y_j according to the probabilities/proportions $p_{ij}(t)$, the probability of going to Y_j from X_i , $i = 1, \dots, n$ and $j = 1, \dots, m$. Here the determination of which recipient state Y_ℓ a particle transitions to depends on which sub-state of X the particle was assigned to upon entering X (see Fig. 5).

Third, in Sect. 3.5.4, the outcome of which recipient state a particle transitions to upon leaving X is determined by a "race" between multiple competing Poisson process k th event time distributions, and is therefore not independent of the time spent in the first state (as in Sect. 3.5.1), nor is it pre-determined upon entry into X (as in Sects. 3.5.2 and 3.5.3). This result is obtained using yet another novel extension of Lemma 1 in which the dwell time in state X is the minimum of $n \geq 2$ independent Poisson process event time distributions.

Lastly (Sect. 3.5.5), we describe an equivalence between (1) the more complex case addressed in Sect. 3.5.4 assuming a dwell time that obeys the minimum of Poisson process 1st event times, before being distributed across m recipient states, and (2) the conceptually simpler case in Sect. 3.5.1 where the dwell time follows a single Poisson process 1st event time distribution before being distributed among m recipient states. This is key to understanding the scope of the Generalized Linear Chain Trick in Sect. 3.7.

3.5.1 Transition to multiple states independent of the X dwell time distribution

Here we extend the case in the previous section and assume that, upon leaving state X, particles can transition to one of m states (call them Y_i , $i = 1, \dots, m$), and that a particle leaving X at time t enters state Y_i with probability $p_i(t)$, where $\sum_{i=1}^m p_i(t) = 1$ [i.e., particles are distributed across all Y_i following a generalized Bernoulli distribution with parameter vector $\mathbf{p}(t) = (p_1(t), \dots, p_m(t))$]. See Fig. 4 for a simple example with constant \mathbf{p} and $m = 2$. An important assumption in this case is that the determination about which state a particle goes to after leaving X is made once it leaves X, and thus the state it transitions to upon exiting X is determined independent of the dwell time in X. Examples from the literature include Model II in Feng et al. (2016), where infected individuals (state X) either recovered (Y_0) or died (Y_1) after an Erlang distributed time delay.

Theorem 4 (Extended LCT with proportional output to multiple states) *Consider the case addressed by Theorem 3, and further assume particles go to one of m states (call them Y_j) with $p_j(t)$ being the probability of going to Y_j . Let S_j be the survival functions for the dwell times in Y_j .*

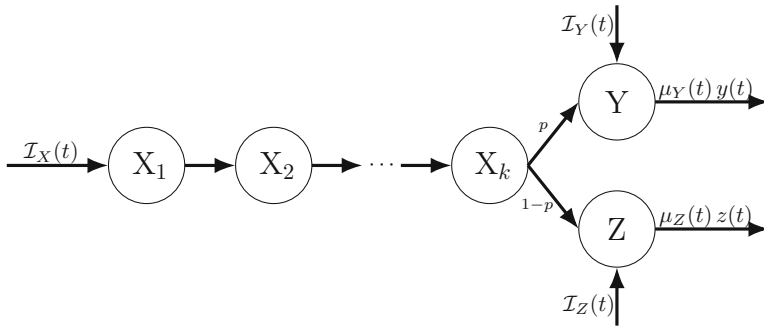


Fig. 4 The special case of Fig. 3 under the assumptions of Theorem 4 (Extended LCT with proportional outputs to multiple states; see Sect. 3.5.1). Specifically, this case assumes that (1) the dwell time distribution for X is Erlang(r, k), and (2) upon exiting X particles are distributed to multiple recipient states, here Y and Z, with probabilities p and $1 - p$, respectively

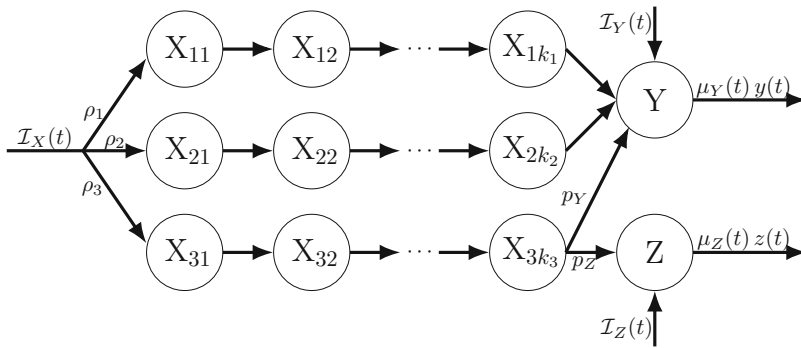


Fig. 5 The sub-state diagram (cf. Fig. 3) for Example 4 after applying Theorem 5 (Extended LCT for finite mixtures of Poisson process event time distributions and output to multiple states). Upon entering X particles have an Erlang(r_i, k_i) dwell time in X with probability $\rho_i, i = 1, 2, 3$, i.e., the X dwell time follows an Erlang mixture distribution (see Sect. 3.5.3)

The mean field integral equations for this case, with $x(0) = x_0$ and $y_j(0) = y_{j0}$, are

$$x(t) = x_0 \mathcal{S}_r^k(t, 0) + \int_0^t \mathcal{I}_X(s) \mathcal{S}_r^k(t, s) ds \tag{38a}$$

$$y_j(t) = y_j(0) \mathcal{S}_j(t, 0) + \int_0^t \left(\mathcal{I}_j(\tau) + p_j(t) \left(x_0 h_r^k(\tau, 0) + \int_0^\tau \mathcal{I}_X(s) h_r^k(\tau, s) ds \right) \right) \mathcal{S}_j(t, \tau) d\tau \tag{38b}$$

These integral equations are equivalent to the following system of equations:

$$\frac{d}{dt} x_1(t) = \mathcal{I}_X(t) - r(t) x_1(t) \tag{39a}$$

$$\frac{d}{dt}x_i(t) = r(t)x_{i-1}(t) - r(t)x_i(t), \quad i = 2, \dots, k \tag{39b}$$

$$y_j(t) = y_j(0)S_j(t, 0) + \int_0^t \left(\mathcal{I}_j(\tau) + p_j(t)r(t)x_k(\tau) \right) S_j(t, \tau) d\tau \tag{39c}$$

where $x(t) = \sum_{i=1}^k x_i(t)$, $x_1(0) = x_0$, $x_i(0) = 0$ for $i \geq 2$, and

$$x_i(t) = x_0 \frac{1}{r(t)} h_r^k(t, 0) + \int_0^t \mathcal{I}_X(s) \frac{1}{r(t)} h_r^k(t, s) ds. \tag{40}$$

Equations (39c) may be further reduced to ODEs, e.g., via Corollary 1.

Proof Equations (39a), (39b) and (40) follow from Theorem 3. Equation (39c) follows from substitution of Eq. (40) into (38b). The derivation of Eq. (38b) is similar to the derivation in Appendix A.1 but accounts for the expected proportion entering each Y_j at time t being equal to $p_j(t)$. \square

Example 3 Consider the example shown in Fig. 4, where the dwell time distribution for X is Erlang(r, k) and the dwell times in Y and Z follow 1st event times under non-homogeneous Poisson processes with respective rates $\mu_Y(t)$ and $\mu_Z(t)$. By Theorem 4 the corresponding mean field ODEs are

$$\frac{d}{dt}x_1(t) = \mathcal{I}_X(t) - r x_1(t) \tag{41a}$$

$$\frac{d}{dt}x_i(t) = r x_{i-1}(t) - r x_i(t), \quad i = 2, \dots, k \tag{41b}$$

$$\frac{d}{dt}y(t) = \mathcal{I}_Y(t) + p r x_k(t) - \mu_Y(t)y(t) \tag{41c}$$

$$\frac{d}{dt}z(t) = \mathcal{I}_Z(t) + (1 - p) r x_k(t) - \mu_Z(t)z(t). \tag{41d}$$

3.5.2 Transition from sub-states of X with differing dwell time distributions and differing output distributions across states Y_j

In this second case, particles in state X can be treated as belonging to a heterogeneous population, where each remains in that state according to one of N possible dwell time distributions, the i th of these being the k_i th event time distribution under a Poisson process with rate $r_i(t)$. Each particle is assigned one of these N dwell time distributions (i.e., it is assigned to sub-state X_i) upon entry into X according to a generalized Bernoulli distribution with a (potentially time varying) probability vector $\rho(t) = (\rho_1(t), \dots, \rho_N(t))$, $\sum_{i=1}^N \rho_i(t) = 1$. In contrast to the previous case, here the outcome of which recipient state a particle transitions to is not necessarily independent of the dwell time distribution.

Note that the dwell time distribution in this case is a finite mixture of N independent Poisson processes event time distributions. If a random variable T is a mixture of Erlang distributions, or more generally a mixture of N independent Poisson process event time distributions, then the corresponding density function (f) and survival function (Φ) are

$$f_{\theta}(t, \tau) = \sum_{i=1}^N \rho_i(\tau) h_{r_i}^{k_i}(t, \tau) \tag{42a}$$

$$\Phi_{\theta}(t, \tau) = \sum_{i=1}^N \rho_i(\tau) S_{r_i}^{k_i}(t, \tau) = \sum_{i=1}^N \rho_i(\tau) \sum_{j=1}^{k_i} \frac{1}{r_i(t)} h_{r_i}^j(t, \tau) \tag{42b}$$

where $\theta(t) = (\rho_1(t), r_1(t), k_1, \dots, \rho_N(t), r_N(t), k_N)$ is the potentially time varying parameter vector for the N distributions that constitute the mixture distribution. Note that if all $r_i(t) = r_i$ are constant, this is a mixture of Erlang distributions, or if also all $k_i = 1$, a mixture of exponentials.

Theorem 5 (Extended LCT for dwell times given by mixtures of Poisson process event time distributions and outputs to multiple states) *Consider a continuous time state transition model with inflow rate $\mathcal{I}_X(t)$ into state X . Assume that the duration of time spent in state X follows a finite mixture of N independent Poisson process event time distributions. That is, X can be partitioned into N sub-states $X_i, i = 1, \dots, N$, each with dwell time distributions given by a Poisson process k_i th event time distributions with rates $r_i(t)$. Suppose the inflow to state X at time t is distributed among this partition according to a generalized Bernoulli distribution with probabilities $\rho_i(t)$, where $\sum_{i=1}^N \rho_i(t) = 1$, so that the input rate to X_i is $\rho_i(t)\mathcal{I}_X(t)$. Assume that particles leaving sub-state X_i then transition to state Y_{ℓ} with probability $p_{i\ell}(t), \ell = 1, \dots, m$, where the duration of time spent in state Y_{ℓ} follows a delay distribution give by survival function S_j . Then we can partition each X_i into $X_{ij}, j = 1, \dots, k_i$, according to Theorem 3 and let $x(t), x_i(t), x_{ij}(t)$, and $y_{\ell}(t)$ be the amounts in states X, X_i, X_{ij} , and Y_{ℓ} at time t , respectively. Assume non-negative initial conditions $x(0) = x_0, x_i(0) = \rho_i(0)x_0, x_{i1}(0) = \rho_i(0)x_0, x_{ij}(0) = 0$ for $j \geq 2$, and $y_{\ell}(0) \geq 0$.*

The mean field integral equations for this scenario are

$$x(t) = x_0 \Phi_{\theta}(t, 0) + \int_0^t \mathcal{I}_X(s) \Phi_{\theta}(t, s) ds \tag{43a}$$

$$y_{\ell}(t) = y_{\ell}(0)S_{\ell}(t, 0) + \int_0^t \left(\mathcal{I}_{\ell}(\tau) + \sum_{i=1}^N p_{ij}(\tau) \left(x_0 \rho_i(\tau) h_{r_i}^{k_i}(\tau, 0) + \int_0^{\tau} \rho_i(s) \mathcal{I}_X(s) h_{r_i}^{k_i}(\tau, s) ds \right) \right) S_{\ell}(t, \tau) d\tau. \tag{43b}$$

The above system of equations (43) are equivalent to

$$\frac{d}{dt}x_{i1}(t) = \rho_i(t)\mathcal{I}_X(t) - r_i(t)x_{i1}(t), \quad i = 1, \dots, N \tag{44a}$$

$$\frac{d}{dt}x_{ij}(t) = r_i(t)(x_{i,j-1}(t) - x_{ij}(t)), \quad i = 1, \dots, N; \quad j = 2, \dots, k_i \tag{44b}$$

$$y_\ell(t) = y_\ell(0)S_\ell(t, 0) + \int_0^t \left(\mathcal{I}_\ell(\tau) + \sum_{i=1}^N r_i(\tau)x_{ik_i}(\tau)p_{i\ell}(\tau) \right) S_\ell(t, \tau) d\tau \tag{44c}$$

with initial conditions $x_{i1}(0) = \rho_i(0)x_0$, $x_{ij}(0) = 0$ for $j \geq 2$, where $x(t) = \sum_{i=1}^N x_i(t)$, and $x_i(t) = \sum_{j=1}^{k_i} x_{ij}(t)$. The amounts in X_i and X_{ij} are

$$x_i(t) = \rho_i(0)x_0\mathcal{S}_{r_i}^{k_i}(t, 0) + \int_0^t \rho_i(s)\mathcal{I}_X(s)\mathcal{S}_{r_i}^{k_i}(t, s)ds \tag{45}$$

$$x_{ij}(t) = \rho_i(0)x_0\frac{h_{r_i}^j(t, 0)}{r_i(t)} + \int_0^t \rho_i(s)\mathcal{I}_X(s)\frac{h_{r_i}^j(t, s)}{r_i(t)}ds. \tag{46}$$

Equations(44c) may be reduced to ODEs, e.g., via Corollary 1.

Proof Substituting Eq.(42b) into Eq.(43a) and then substituting Eq.(45) yields $x(t) = \sum_{i=1}^N x_i(t)$. Applying Theorem 3 to each X_i [i.e., to each Eq. (45)] then yields Eqs. (46), (44a) and (44b). (Alternatively, one could prove this directly by differentiating Eq. (46) using Eq. (30) from Lemma 2). The $y_\ell(t)$ equations (44c) are obtained from (43b) by substitution of Eq. (46). □

Example 4 Consider the scenario in Fig. 5, where particles entering state X at rate $\mathcal{I}_X(t)$ enter sub-state X_i with probability ρ_i , $\rho_1 + \rho_2 + \rho_3 = 1$, and the X_i dwell time is Erlang(r_i, k_i) distributed. Particles exiting X_1 and X_2 transition to Y with probability 1, while particles exiting X_3 transition either to state Y or Z with probabilities p_Y and $p_Z = 1 - p_Y$. Assume particle may also enter states Y and Z from sources other than state X (at rates $\mathcal{I}_Y(t)$ and $\mathcal{I}_Z(t)$, respectively), and the dwell times in those two states follow the 1st event times of independent nonhomogeneous Poisson processes with rates $\mu_Y(t)$ and $\mu_Z(t)$. Theorem 5 yields the following mean field ODEs (see Fig. 5).

$$\frac{d}{dt}x_{i,1}(t) = \rho_i\mathcal{I}_X(t) - r_i x_{i,1}(t), \quad i = 1, \dots, 3, \tag{47a}$$

$$\frac{d}{dt}x_{i,j}(t) = r_i(x_{i,j-1}(t) - x_{ij}(t)), \quad j = 2, \dots, k_i \tag{47b}$$

$$\begin{aligned} \frac{d}{dt}y(t) &= \mathcal{I}_Y(t) + r_1 x_{1,k_1}(t) + r_2 x_{2,k_2}(t) \\ &\quad + r_3 p_Y x_{3,k_3}(t) - \mu_Y(t)y(t) \end{aligned} \tag{47c}$$

$$\frac{d}{dt}z(t) = \mathcal{I}_Z(t) + r_3 p_Z x_{3,k_3}(t) - \mu_Z(t)z(t). \tag{47d}$$

3.5.3 Extended LCT for dwell times given by finite mixtures of Poisson process event time distributions

It's worth noting that in some applied contexts one may want to approximate a non-Erlang delay distribution with a mixture of Erlang distributions (see Sect. 3.7.1 and Appendix B for more details on making such approximations). Theorem 5 above details how assuming such a mixture distribution would be reflected in the structure of the corresponding mean field ODEs. This case can also be addressed in the more general context provided in Sect. 3.7.1.

3.5.4 Transition to multiple states following "competing" Poisson processes

We now consider the case where T , the time a particle spends in a given state X , follows the distribution given by $T = \min_i T_i$, the minimum of $n \geq 2$ independent random variables T_i , where T_i has either an Erlang(r_i, k_i) distribution or, more generally, a Poisson process k_i th event time distribution with rate $r_i(t)$. Upon leaving state X , particles have the possibility of transitioning to any of m recipient states Y_ℓ , $\ell = 1, \dots, m$, where the probability of transitioning to state Y_ℓ depends on which of the n random variables T_i was the minimum. That is, if a particle leaves X at time $T = T_i = t$, then the probability of entering state Y_ℓ is $p_{i\ell}(t)$.

The distribution associated with T is not itself an Erlang distribution or a Poisson process event time distribution, however its survival function is the product⁷ of such survival functions, i.e.,

$$\mathcal{S}(t, \tau) \equiv \prod_{i=1}^n \mathcal{S}_{r_i}^{k_i}(t, \tau). \quad (48)$$

As detailed below, we can further generalize the recursion relation in Lemma 1 for the distributions just described above, which can then be used to produce a mean field system of ODEs based on appropriately partitioning X into sub-states.

Before considering this case in general, it is helpful to first describe the sub-states of X imposed by assuming the dwell time distribution described above, particularly the case where the distribution for each T_i is based on 1st event times (i.e., all $k_i = 1$). Recall that the minimum of n exponential random variables (which we may think of as 1st event times under a homogeneous Poisson process) is exponential with a rate that is the sum of the individual rates $r = \sum_{i=1}^n r_i$. More generally, it is true that the minimum of n 1st event times under independent Poisson processes with rates $r_i(t)$ is itself distributed as the 1st event time under a single Poisson process with rate $r(t) \equiv \sum_{i=1}^n r_i(t)$, thus in this case $\mathcal{S}(t, \tau) = \prod_{i=0}^n \mathcal{S}_{r_i}^1(t, \tau) = \mathcal{S}_r^1(t, \tau)$. Additionally, if particles leaving state X are then distributed across the recipient states Y_ℓ as described above, then this scenario is equivalent to the proportional outputs case described in Theorem 4 with a dwell time that follows a Poisson process 1st event time distribution with rate $r(t) \equiv \sum_{i=1}^n r_i(t)$ and a probability vector $p_\ell =$

⁷ It is generally true that the survival function for a minimum of multiple independent random variables is the product of their survival functions.

$\sum_{i=1}^n p_{i\ell}(t)r_i(t)/r(t)$, since $P(T = T_i) = r_i(T)/r(T)$. (This mean field equivalence of these two cases is detailed in Sect. 3.5.5.) Thus, the natural partitioning of X in this case is into sub-states with dwell times that follow *iid* 1st event time distributions with rate $r(t) \equiv \sum_{i=1}^N r_i(t)$.

We may now describe the mean field ODEs for the more general case using the following notation. To index the sub-states of X , consider the i th Poisson process and its k_i th event time distribution which defines the distribution of T_i . Let $a_i \in \{1, \dots, k_i\}$ denote the event number a particle is awaiting under the i th Poisson process. Then we can describe the particle's progress through X according to its progress along each of these n Poisson processes using the index vector $\alpha \in \mathcal{K}$, where

$$\mathcal{K} = \{(a_1, a_2, \dots, a_n) \mid a_j \in \{1, \dots, k_j\}\}. \tag{49}$$

Let $\mathcal{K}_i \subset \mathcal{K}$ denote the subset of indices where $a_i = k_i$ (where we think of particles in these sub-states as being poised to reach the k_i th event related to the i th Poisson process, and thus poised to transition out of state X).

To extend Lemma 2 for these distributions, define

$$u(t, \tau, \alpha) \equiv \prod_{i=1}^n e^{-m_i(t,\tau)} \frac{m_i(t, \tau)^{a_i-1}}{(a_i - 1)!} \tag{50}$$

where $m_i(t, \tau) = \exp(-\int_{\tau}^t r_i(s)ds)$, and $u(\tau, \tau, \alpha) = 1$ if $\alpha = (1, \dots, 1)$ and $u(\tau, \tau, \alpha) = 0$ otherwise. Note that $\prod_{i=1}^n h_{r_i}^{a_i}(t, \tau) = u(t, \tau, \alpha) \prod_{i=1}^n r_i(t)$ (c.f. Lemma 2). Applying Eq. (9b) to $\mathcal{S}(t, \tau)$ in Eq. (48), the survival function given by Eq. (48) [c.f. Eq. (31) and (9b)] can be written

$$\mathcal{S}(t, \tau) = \sum_{\alpha \in \mathcal{K}} u(t, \tau, \alpha). \tag{51}$$

We will also refer to the quantities u and \mathcal{S} with the j th element of each product (in u) removed using the notation

$$u_{\setminus j}(t, \tau, \alpha) \equiv \prod_{i=1, i \neq j}^n e^{-m_i(t,\tau)} \frac{m_i(t, \tau)^{a_i-1}}{(a_i - 1)!} \tag{52a}$$

$$\mathcal{S}_{\setminus j}(t, \tau) \equiv \sum_{\alpha \in \mathcal{K}_j} u_{\setminus j}(t, \tau, \alpha). \tag{52b}$$

This brings us to the following lemma, which generalizes Lemma 1 and Lemma 2 to distributions that are the minimum of n different (independent) Poisson process event times. As with the above lemmas, Lemma 3 will allow one to partition X into sub-states corresponding to each of the event indices in \mathcal{K} describing the various stages of progress along each Poisson process prior to the first of them reaching the target event number.

Lemma 3 For u as defined in Eq. (50), differentiation with respect to t yields

$$\frac{d}{dt}u(t, \tau, \alpha) = \sum_{j=1}^n r_j(t) u(t, \tau, \alpha_{j,-1}) \mathbb{1}_{[a_j>1]}(\alpha) - \sum_{j=1}^n r_j(t) u(t, \tau, \alpha) \quad (53)$$

where the notation $\alpha_{j,-1}$ denotes the index vector generated by decrementing the j th element of α , a_j [assuming $a_j > 1$; for example, $\alpha_{2,-1} = (a_1, a_2 - 1, \dots, a_n)$], and the indicator function $\mathbb{1}_{[a_j>1]}(\alpha)$ is 1 if $a_j > 1$ and 0 otherwise.

Proof Using the definition of u in Eq. (50) above, it follows that

$$\begin{aligned} \frac{d}{dt}u(t, \tau, \alpha) &= \frac{d}{dt} \prod_{i=1}^n e^{-m_i(t,\tau)} \frac{m_i(t, \tau)^{a_i-1}}{(a_i - 1)!} \\ &= \sum_{j=1}^n \left(\prod_{\substack{i=1 \\ i \neq j}}^n e^{-m_i(t,\tau)} \frac{m_i(t, \tau)^{a_i-1}}{(a_i - 1)!} \right) \left[-r_j(t) e^{-m_j(t,\tau)} \frac{m_j(t, \tau)^{a_j-1}}{(a_j - 1)!} \right. \\ &\quad \left. + \mathbb{1}_{[a_j>1]}(\alpha) r_j(t) e^{-m_j(t,\tau)} \frac{m_j(t, \tau)^{a_j-2}}{(a_j - 2)!} \right] \\ &= \sum_{j=1}^n -r_j(t) \prod_{i=1}^n e^{-m_i(t,\tau)} \frac{m_i(t, \tau)^{a_i-1}}{(a_i - 1)!} \\ &\quad + \sum_{j=1}^n \mathbb{1}_{[a_j>1]}(\alpha) r_j(t) e^{-m_j(t,\tau)} \frac{m_j(t, \tau)^{a_j-2}}{(a_j - 2)!} \prod_{\substack{i=1 \\ i \neq j}}^n e^{-m_i(t,\tau)} \frac{m_i(t, \tau)^{a_i-1}}{(a_i - 1)!} \\ &= \sum_{j=1}^n r_j(t) u(t, \tau, \alpha_{j,-1}) \mathbb{1}_{[a_j>1]}(\alpha) - \sum_{j=1}^n r_j(t) u(t, \tau, \alpha). \end{aligned} \quad (54)$$

□

The next theorem details the LCT extension that follows from Lemma 3.

Theorem 6 (Extended LCT for dwell times given by competing Poisson processes) Consider a continuous time dynamical system model of mass transitioning among multiple states, with inflow rate $\mathcal{I}_X(t)$ to a state X . The distribution of time spent in state X (call it T) is the minimum of n random variables, i.e., $T = \min_i(T_i)$, $i = 1, \dots, n$, where T_i are either Erlang(r_i, k_i) distributed or follow the more general (nonhomogeneous) Poisson process k_i th event time distribution with rate $r_i(t)$. Assume particles leaving X can enter one of m states Y_ℓ , $\ell = 1, \dots, m$. If a particle leaves X at time T_i (i.e., T_i occurred first, so $T = T_i$), and then the particle transitions into state Y_ℓ with probability $p_{i\ell}(T)$. Let $x(t)$, and $y_\ell(t)$ be the amount in each state, respectively, at time t , and assume non-negative initial conditions.

The mean field integral equations for this scenario, for $\ell = 1, \dots, m$ and $i = 1, \dots, n$, are

$$x(t) = x_0 \mathcal{S}(t, 0) + \int_0^t \mathcal{I}_X(s) \mathcal{S}(t, s) ds \tag{55a}$$

$$y_\ell(t) = y_\ell(0)S_\ell(t, 0) + \int_0^t \left(\mathcal{I}_\ell(\tau) + \sum_{i=1}^n p_{i\ell} \left(x_0 \mathcal{S}_{\setminus i}(\tau, 0) h_{r_i}^{k_i}(\tau, 0) + \int_0^\tau \mathcal{I}_X(s) \mathcal{S}_{\setminus i}(\tau, s) h_{r_i}^{k_i}(\tau, s) ds \right) \right) S_\ell(t, \tau) d\tau. \tag{55b}$$

Equations (55) above are equivalent to

$$\frac{d}{dt} x_{(1, \dots, 1)}(t) = \mathcal{I}_X(t) - r(t) x_{(1, \dots, 1)}(t), \tag{56a}$$

$$\frac{d}{dt} x_\alpha(t) = \sum_{i=1}^n r_i(t) x_{\alpha_{i,-1}}(t) \mathbb{1}_{[a_i > 1]}(\alpha) - r(t) x_\alpha(t) \tag{56b}$$

$$y_\ell(t) = y_\ell(0)S_\ell(t, 0) + \int_0^t \left(\mathcal{I}_\ell(\tau) + \sum_{i=1}^n p_{i\ell}(\tau) \sum_{\alpha \in \mathcal{K}_i} r_i(t) x_\alpha(\tau) \right) S_\ell(t, \tau) d\tau \tag{56c}$$

for all $\alpha \in \mathcal{K} \setminus (1, \dots, 1)$, $r(t) = \sum_{i=1}^n r_i(t)$, $x(t) = \sum_{\alpha \in \mathcal{K}} x_\alpha(t)$, and

$$x_\alpha(t) = x_0 u(t, 0, \alpha) + \int_0^t \mathcal{I}_X(s) u(t, s, \alpha) ds. \tag{57}$$

The $y_\ell(t)$ equations (56c) may be further reduced to a system of ODEs, e.g., via Corollary 1.

Proof Substituting Eq. (51) into Eq. (55a) yields

$$x(t) = x_0 \sum_{\alpha \in \mathcal{K}} u(t, 0, \alpha) + \int_0^t \mathcal{I}_X(s) \sum_{\alpha \in \mathcal{K}} u(t, s, \alpha) ds \tag{58}$$

$$= \sum_{\alpha \in \mathcal{K}} \left(x_0 u(t, 0, \alpha) + \int_0^t \mathcal{I}_X(s) u(t, s, \alpha) ds \right) = \sum_{\alpha \in \mathcal{K}} x_\alpha(t).$$

Differentiating (57) yields equations Eqs. (56a) and (56b) as follows. First, if $\alpha = (1, \dots, 1)$ then by Lemma 3

$$\begin{aligned} \frac{d}{dt} x_{(1, \dots, 1)}(t) &= -x_0 \sum_{i=1}^n r_i(t) u(t, 0, \alpha) \\ &\quad - \sum_{i=1}^n r_i(t) \int_0^t \mathcal{I}_X(s) u(t, s, \alpha) ds + \mathcal{I}_X(t) \\ &= \mathcal{I}_X(t) - \sum_{i=1}^n r_i(t) x_{(1, \dots, 1)}(t). \end{aligned} \tag{59}$$

Next, if α has any $a_i > 1$, differentiating Eq. (57) and applying Lemma 3 yields

$$\begin{aligned} \frac{d}{dt}x_\alpha(t) &= x_0 \frac{d}{dt}u(t, 0, \alpha) + \int_0^t \mathcal{I}_X(s) \frac{d}{dt}u(t, s, \alpha) ds \\ &= x_0 \left(\sum_{i=1}^n r_i(t)u(t, 0, \alpha_{i,-1}) \mathbb{1}_{[a_i>1]}(\alpha) - \sum_{i=1}^n r_i(t)u(t, \alpha) \right) \\ &\quad + \int_0^t \mathcal{I}_X(s) \left(\sum_{i=1}^n r_i(t)u(t, s, \alpha_{i,-1}) \mathbb{1}_{[a_i>1]}(\alpha) - \sum_{i=1}^n r_i(t)u(t, s, \alpha) \right) ds \\ &= \sum_{i=1}^n r_i(t) x_{\alpha_{i,-1}}(t) \mathbb{1}_{[a_i>1]}(\alpha) - \sum_{i=1}^n r_i(t) x_\alpha(t) \end{aligned} \tag{60}$$

Note that, by the definitions of x_α and u that initial condition $x(0) = x_0$ becomes $x_{(1,\dots,1)}(0) = x_0$ and $x_\alpha(0) = 0$ for the remaining $\alpha \in \mathcal{K}$.

Eqs. (55b) become (56c), where $\mathcal{K}_i = \{\alpha \mid \alpha \in \mathcal{K}, a_i = k_i\}$, by substituting Eqs. (52), (57), and $\mathcal{S}_{\setminus i}(t, \tau) h_{r_i}^{k_i}(t, \tau) = \sum_{\alpha \in \mathcal{K}_i} r_i(t) u(t, \tau, \alpha)$, which yields

$$\begin{aligned} x_0 \mathcal{S}_{\setminus i}(\tau, 0) h_{r_i}^{k_i}(\tau, 0) &+ \int_0^\tau \mathcal{I}_X(s) \mathcal{S}_{\setminus i}(\tau, s) h_{r_i}^{k_i}(\tau, s) ds \\ &= x_0 \sum_{\alpha \in \mathcal{K}_i} r_i(t) u(\tau, 0, \alpha) + \int_0^\tau \mathcal{I}_X(s) \sum_{\alpha \in \mathcal{K}_i} r_i(t) u(\tau, s, \alpha) ds \\ &= r_i(t) \sum_{\alpha \in \mathcal{K}_i} \left(x_0 u(\tau, 0, \alpha) + \int_0^\tau \mathcal{I}_X(s) u(\tau, s, \alpha) ds \right) = \sum_{\alpha \in \mathcal{K}_i} r_i(t) x_\alpha(\tau). \end{aligned} \tag{61}$$

□

Example 5 See Fig. 6. Suppose the X dwell time is $T = \min(T_1, T_2)$ where T_1 and T_2 are the k_1 th and k_2 th event time distributions under independent Poisson processes (call these PP1 and PP2) with rates $r_1(t)$ and $r_2(t)$, respectively. Assume that, upon leaving X, particles transition to Y_1 if $T = T_1$ or to Y_2 if $T = T_2$. By Theorem 6, we can partition X into sub-states defined by which event (under each Poisson process) particles are awaiting next. Upon entry into X, all particles enter sub-state $X_{1,1}$ where they each await the 1st events under PP1 or PP2. If the next event to occur for a given particle is from PP1, the particle transitions to $X_{2,1}$ where it awaits a 2nd event from PP1 or 1st event from PP2 (hence the subscript notation). Likewise, if PP2’s 1st event occurs before PP1’s 1st event, the particle would transition to $X_{1,2}$, and so on. Particles would leave these two states to either $X_{2,2}$, Y_1 , or Y_2 depending on which event occurs next. Under these assumptions, with $k_1 = k_2 = 2$ and exponential Y_i dwell times with rates μ , then the corresponding mean field equations (using $r(t) = r_1(t) + r_2(t)$) are

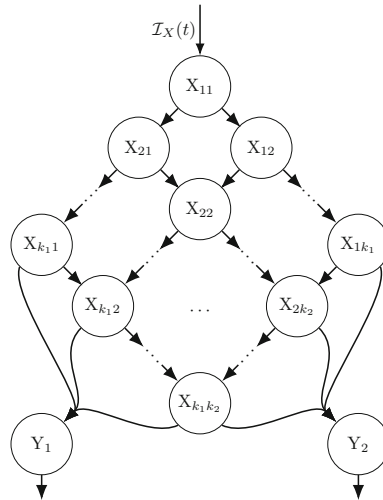


Fig. 6 The sub-state diagram (cf. Fig. 3) for the scenario detailed in Example 5 after the application of Theorem 6 (extended LCT for dwell times given by competing Poisson processes). Here the X dwell time distribution is the minimum of two Erlang random variables $T_i \sim \text{Erlang}(r_i, k_i)$, $i = 1, 2$, which can be thought of as event time distribution under two homogeneous Poisson processes as detailed in the main text. We here assume that whichever of these occurs first determines whether particles leaving X transition to Y_1 or Y_2 , respectively

$$\frac{dx_{11}}{dt} = \mathcal{I}_X(t) - r(t) x_{11}(t) \tag{62a}$$

$$\frac{dx_{21}}{dt} = r_1(t) x_{11}(t) - r(t) x_{21}(t) \tag{62b}$$

$$\frac{dx_{12}}{dt} = r_2(t) x_{11}(t) - r(t) x_{12}(t) \tag{62c}$$

$$\frac{dx_{22}}{dt} = r_1(t) x_{12}(t) + r_2(t) x_{21}(t) - r(t) x_{22}(t) \tag{62d}$$

$$\frac{dy_1}{dt} = r_1(t) x_{22}(t) - \mu(t) y_1(t) \tag{62e}$$

$$\frac{dy_2}{dt} = r_2(t) x_{22}(t) - \mu(t) y_2(t). \tag{62f}$$

It's worth pointing out that the dwell times for the above sub-states of X are all identically distributed Poisson process 1st event times (note the loss rates in Eqs. (62a)–(62d), and recall the weak memorylessness property from Sect. 3.1.2). All particles in a X sub-state at time τ will spend a remaining amount of time in that state that follows a 1st event time distributions under a Poisson process with rate $r(t) = r_1(t) + r_2(t)$. This is a slight generalization of the familiar fact that the minimum of n independent exponentially distributed random variables (with respective rates r_i) is itself an exponential random variable (with rate $r \equiv \sum_{i=1}^n r_i$). The next section addresses the generality of this observation about the sub-states of X.

3.5.5 Mean field equivalence of proportional outputs and competing Poisson processes for 1st event time distributions

The scenarios described in Sect. 3.5.1 (proportional distribution across multiple states Y_ℓ after an Erlang dwell time in X) and Sect. 3.5.4 (proportional distribution across multiple states based upon competing Poisson processes), can lead to equivalent mean field equations when the X dwell times follow Poisson process 1st event time distributions, as is Example 5. This equivalence is detailed in Theorem 7, and is an important aspect of the GLCT detailed in Sect. 3.7 because it helps to show how sub-states with dwell times distributed as Poisson process 1st event times are the fundamental buildings blocks of the GLCT.

Theorem 7 (Mean field equivalence of proportional outputs and competing Poisson processes for 1st event time distributions) *Consider the special case of Theorem 6 (the Extended LCT for competing Poisson processes) where X has a dwell time given by $T = \min_i T_i$, where each T_i is a Poisson process 1st event time with rate $r_i(t)$, $i = 1, \dots, n$ and particles transition to Y_ℓ with probability $p_{i\ell}(T)$ when $T = T_i$. The corresponding mean field model is equivalent to the special case of Theorem 4 (the Extended LCT for multiple outputs) where the X dwell time is a Poisson process 1st event time distribution with rate $r(t) = \sum_{i=1}^n r_i(t)$, and the transition probability vector for leaving X and entering state Y_ℓ is given by $p_\ell(t) = \sum_{i=1}^n p_{i\ell}(t) r_i(t)/r(t)$.*

Proof First, in this case $\mathcal{S}(t, \tau) = \prod_{i=0}^n \mathcal{S}_{r_i}^1(t, \tau) = \mathcal{S}_r^1(t, \tau)$. Since all $k_i = 1$, the probability that $T = T_i$ is $r_i(T)/r(T)$, thus the probability that a particle leaving X at t goes to Y_ℓ is $p_\ell(t) = \sum_{i=1}^n \frac{r_i(t)}{r(t)} p_{i\ell}(t)$. Substituting the above equalities into the mean field Eq. (56a) (where there's only one possible index in $\mathcal{K} = \{(1, 1, \dots, 1)\}$) and (56c) gives

$$\frac{d}{dt}x(t) = \mathcal{I}_X(t) - r(t)x(t) \tag{63a}$$

$$y_j(t) = y_j(0)S_j(t, 0) + \int_0^t \left(\mathcal{I}_j(\tau) + r(t)p_j(\tau)x(\tau) \right) S_j(t, \tau) d\tau \tag{63b}$$

which are the mean field equations for the aforementioned special case of Theorem 4. □

3.6 Modeling intermediate state transitions: reset the clock, or not?

We next describe how to apply extensions of the LCT in two similar but distinctly different scenarios (see Fig. 7) where the transition to one or more intermediate sub-states either resets an individual's overall dwell time in state X (by assuming the time spent in an intermediate sub-state X_{I_i} is independent of time already spent in X_0 ; see Sect. 3.6.1), or instead leaves the overall dwell time distribution for X unchanged (by conditioning the time spent in intermediate state X_{I_i} on the time already spent in X_0 ; see Sect. 3.6.2).

An example of these different assumptions leading to important differences in practice comes from Feng et al. (2016) where individuals infected with Ebola can either

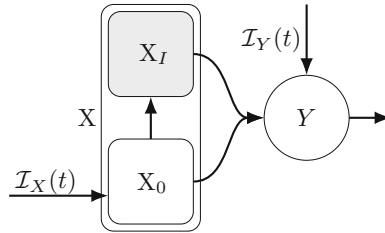


Fig. 7 Should the overall dwell time distribution for state X be “reset” by the transition from base sub-state X_0 to intermediate sub-state X_I (i.e., should the dwell time in state X_I be independent of the time already spent in X_0 ?), or should the X_I dwell time be conditioned on time already spent in X_0 so that the $X_0 \rightarrow X_I$ transition does not alter the overall dwell time in state X? How do these different assumptions alter the structure of the corresponding mean field ODEs? We answer these question in Sect. 3.6 where we describe how to apply the LCT in scenarios with intermediate states, assuming in Sect. 3.6.1 that the dwell time distribution for X_I is independent of the amount of time spent in X_0 , and assuming in Sect. 3.6.2 that the overall dwell time for X is unaffected by transitions from X_0 to X_I

leave the infected state (X) directly (either to a recovery or death), or after first transitioning to an intermediate hospitalized state (X_I) which needs to be explicitly modeled in order to incorporate a quarantine effect into the rates of disease transmission (i.e., the force of infection should depend on the number of non-quarantined individuals, i.e., X_0). As shown in Feng et al. (2016), the epidemic model output depends strongly upon whether or not it is assumed that moving into the hospitalized sub-state impacts the distribution of time spent in the infected state X.

To most simply illustrate these two scenarios, consider the simple case in Fig. 7 where a single intermediate sub-state X_I is being modeled, and particles enter X into sub-state X_0 at rate $\mathcal{I}_X(t)$. Let $X = X_0 \cup X_I$. Both cases assume particles transition out of X_0 either to sub-state X_I or they leave state X directly and enter state Y. Both cases also assume the distribution of time spent in X_0 is $T_* = \min(T_0, T_1)$ where particles transition to X_I if $T_1 < T_0$ (i.e., if $T = T_1$) or to Y if $T_0 < T_1$ (where each T_i is the k_i th event time under Poisson processes with rates $r_0(t)$ and $r_1(t)$ (see Sects. 3.5.4 and 3.5.5). Let T_I denote the distribution of time spent in intermediate state X_I . The first case assumes T_I is independent of time spent in X_0 (i.e., the transition to X_I ‘resets the clock’; see Sect. 3.6.1). The second case assumes T_I is conditional on time already spent in X_0 (call it t_0), such that the total amount of time spent in X, $t_0 + T_I$, is equivalent in distribution to T_0 (i.e., the transition to X_I does not change the overall distribution of time spent in X; see Sect. 3.6.2).

In the next two sections, we provide extensions of the LCT for generalizations of these two scenarios, extended to multiple possible intermediate states with eventual transitions out of X into multiple recipient states.

3.6.1 Intermediate states that reset dwell time distributions

First, we consider the case in which the time spent in the intermediate state X_I is independent of the time already spent in X (i.e., in the base state X_0). This is arguably the more commonly encountered (implicit) assumption found in ODE models that

aren't explicitly derived from a stochastic model and/or mean field integro-differential delay equations.

The construction of mean field ODEs for this case is a straightforward application of Theorem 6 from the previous section, combined with the extended LCT with output to multiple states (Theorem 4). Here we have extended this scenario to include M_X intermediate sub-states X_{I_j} where the transition to those sub-states from base state X_0 is based on the outcome of N competing Poisson process event time distributions (T_i), and upon leaving the intermediate states particles transition out of state X into one of M_Y possible recipient states Y_ℓ .

Theorem 8 (Extended LCT with dwell time altering intermediate sub-state transitions) *Suppose particles enter X at rate $\mathcal{I}_X(t)$ into a base sub-state X_0 . Assume particles remain in X_0 according to a dwell time distribution given by T , the minimum of $N + 1$ independent Poisson process k_i th event time distributions with rates $r_i(t)$, $i = 0, \dots, N$, i.e., $T = \min_i(T_i)$. Particles leaving X_0 transition to one of $M_X \geq 1$ intermediate sub-states X_{I_i} or to one of $M_Y \geq 1$ recipient states Y_ℓ according to which $T_i = T$. If $T_0 = T$ then the particle leaves X and the probability of transitioning to Y_ℓ is $p_{0\ell}(T)$, where $\sum_{\ell=1}^{M_Y} p_{0\ell}(T) = 1$. If $T_i = T$ for $i \geq 1$ then the particle transitions to X_{I_j} with probability $p_{ij}(T)$, where $\sum_{j=1}^{M_X} p_{ij}(T) = 1$. Particles in intermediate state X_{I_j} remain there according to the κ_i th event times under a Poisson process with rate $\varrho_i(t)$, and then transition to state Y_ℓ with probability $q_{j\ell}(t)$, where (for fixed t) $\sum_{\ell=1}^{M_Y} q_{j\ell}(t) = 1$, and they remain in Y_ℓ according to a dwell time with survival function $S_\ell(t, \tau)$.*

In this case the corresponding mean field equations are

$$\frac{d}{dt}x_{0(1,\dots,1)}(t) = \mathcal{I}_X(t) - \sum_{i=0}^N r_i(t) x_{(1,\dots,1)}(t) \tag{64a}$$

$$\frac{d}{dt}x_{0\alpha}(t) = \sum_{i=0}^N r_i(t) \left(x_{0\alpha_{i-1}}(t) \mathbb{1}_{[a_i > 1]}(\alpha) - x_{0\alpha}(t) \right) \tag{64b}$$

$$\frac{d}{dt}x_{I_{j1}}(t) = \mathcal{I}_{X_{I_j}}(t) + p_{ij}(t) \left(\sum_{\alpha \in \mathcal{K}_i} r_i(t) x_{0\alpha}(t) \right) - \varrho_j(t) x_{I_{j1}}(t) \tag{64c}$$

$$\frac{d}{dt}x_{I_{jk}}(t) = \varrho_j(t) (x_{I_{j,k-1}}(t) - x_{I_{jk}}(t)), \quad k = 2, \dots, \kappa_j \tag{64d}$$

$$y_\ell(t) = y_\ell(0) S_\ell(t, 0) + \int_0^t \left(\mathcal{I}_{Y_\ell}(\tau) + p_{0\ell}(\tau) \sum_{\alpha \in \mathcal{K}_0} r_0(\tau) x_{0\alpha}(\tau) + \sum_{j=1}^{M_X} \varrho_j(\tau) x_{I_{j\kappa_j}}(\tau) q_{j\ell}(\tau) \right) S_Y(t, \tau) d\tau. \tag{64e}$$

where the amount in base sub-state X_0 is $x_0(t) = \sum_{\alpha \in \mathcal{K}} x_{0\alpha}(t)$, the amount in the j th intermediate state X_{I_j} is $x_{I_j}(t) = \sum_{k=1}^{\kappa_j} x_{I_{jk}}(t)$ (see Theorem 6 for notation), $\mathcal{K} = \{(a_0, a_1, \dots, a_N) \mid a_j \in \{1, \dots, \kappa_j\}\}$, $\mathcal{K}_i = \{\alpha \in \mathcal{K} \mid a_i = \kappa_i\}$, $j = 1, \dots, N$,

$\ell = 1, \dots, M_Y$, and in Eq. (64b) $\alpha = (a_0, \dots, a_N) \in \mathcal{K} \setminus (1, \dots, 1)$. Note that the $y_\ell(t)$ equations (64e) may be further reduced to a system of ODEs, e.g., via Corollary 1, and that more complicated distributions for dwell times in intermediate states X_{I_i} (e.g., an Erlang mixture) could be similarly modeled according to other cases addressed in this manuscript.

Proof This follows from applying Theorem 6 to X_0 and treating the intermediate states X_{I_j} as recipient states, then applying Theorem 4 to each intermediate state to partition each X_{I_j} into $X_{I_{jk}}$, $k = 1, \dots, \kappa_j$, yielding Eq. (64). \square

Example 6 Consider the scenario in Fig. 7. Let the X_0 dwell time be the minimum of $T_0 \sim \text{Erlang}(r_0, 2)$ and $T_1 \sim \text{Erlang}(r_1, 2)$, with intermediate state dwell time $T_{I_1} \sim \text{Erlang}(\varrho_1, \kappa_1 = 3)$ and an exponential (rate μ) dwell time in Y . Assume the only inputs into X are into X_0 at rate $\mathcal{I}_X(t)$. By Theorem 8 the corresponding mean field ODEs (see Fig. 8) are Eq. (65), where $x_0(t) = x_{0(1,1)}(t) + x_{0(2,1)}(t) + x_{0(1,2)}(t) + x_{0(2,2)}(t)$ and $x_{I_1}(t) = x_{I_{11}}(t) + x_{I_{12}}(t) + x_{I_{13}}(t)$.

$$\frac{d}{dt}x_{0(1,1)}(t) = \mathcal{I}_X(t) - (r_0 + r_1)x_{0(1,1)}(t) \tag{65a}$$

$$\frac{d}{dt}x_{0(2,1)}(t) = r_0x_{0(1,1)}(t) - (r_0 + r_1)x_{0(2,1)}(t) \tag{65b}$$

$$\frac{d}{dt}x_{0(1,2)}(t) = r_1x_{0(1,1)}(t) - (r_0 + r_1)x_{0(1,2)}(t) \tag{65c}$$

$$\frac{d}{dt}x_{0(2,2)}(t) = r_0x_{0(1,2)}(t) + r_1x_{0(2,1)}(t) - (r_0 + r_1)x_{0(2,2)}(t) \tag{65d}$$

$$\frac{d}{dt}x_{I_{11}}(t) = r_1x_{0(1,2)}(t) + r_1x_{0(2,2)}(t) - \varrho_1x_{I_{11}}(t) \tag{65e}$$

$$\frac{d}{dt}x_{I_{12}}(t) = \varrho_1x_{I_{11}}(t) - \varrho_1x_{I_{12}}(t) \tag{65f}$$

$$\frac{d}{dt}x_{I_{13}}(t) = \varrho_1x_{I_{12}}(t) - \varrho_1x_{I_{13}}(t) \tag{65g}$$

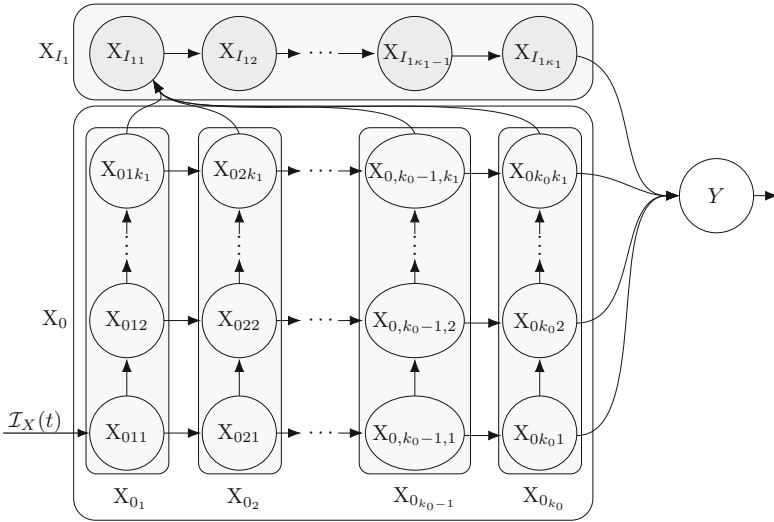
$$\frac{d}{dt}y(t) = \mathcal{I}_Y(t) + r_0x_{0(2,1)}(t) + r_0x_{0(2,2)}(t) + \varrho_1x_{I_{13}}(t) - \mu y(t). \tag{65h}$$

3.6.2 Intermediate states that preserve dwell time distributions

In this section we address how to modify the outcome in the previous section to instead construct mean field ODE models that incorporate ‘dwell time neutral’ sub-state transitions, i.e., where the distribution of time spent in X is the same regardless of whether or not particles transition (within X) from some base sub-state X_0 to one or more intermediate sub-states X_{I_j} . This is done by conditioning the dwell time distributions in X_{I_i} on time spent in X_0 in a way that leverages the weak memorylessness property discussed in Sect. 3.1.2.

In applications, this case (in contrast to the previous case) is perhaps the more commonly desired assumption, since modelers often seek to partition states into sub-states where key characteristics, like the overall X dwell time distribution, remain

(a) Intermediate state dwell time is independent of time spent in X_0 .



(b) Intermediate state dwell time is conditioned on progress through X_0 .

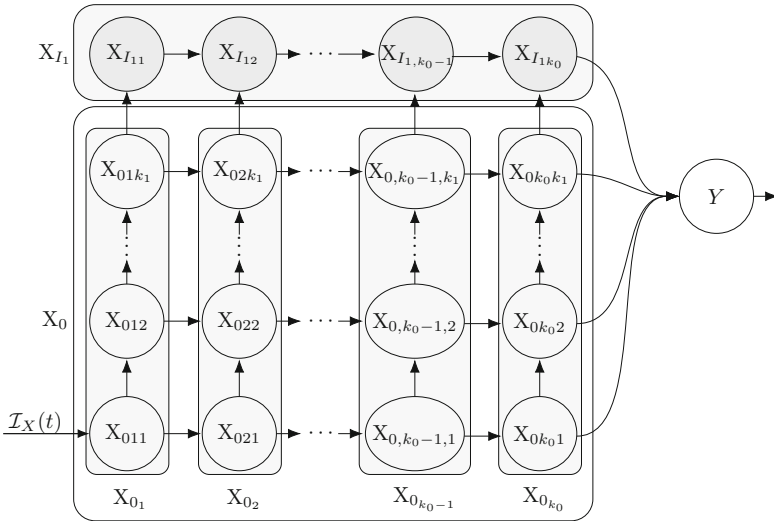


Fig. 8 When modeling sub-state transitions, e.g., from a base state X_0 to intermediate state X_{I_1} (state $X = X_0 \cup X_{I_1}$), the intermediate sub-state dwell time can be (a) independent of time already spent in X_0 (Sect. 3.6.1, Theorem 8), or (b) conditioned on that time so that the dwell time in state X is unaffected by the sub-state transition (Sect. 3.6.2, Theorem 9). In both cases, the dwell time distribution for X_0 is the minimum of independent Erlang distributions, and sub-states within X_0 are as discussed in Sect. 3.5.4. **Panel (a):** The dwell time in X_{I_1} is Erlang(ϱ_1, κ_1) and independent of time spent in X_0 . Thus, the sub-state transition alters the overall dwell time in state X . For the more general case, see Sect. 3.6.1 and Theorem 8. **Panel (b):** The overall X dwell time distribution $T_0 \sim \text{Erlang}(r_0, k_0)$ is preserved by conditioning the dwell time in the intermediate state on the prior progress through sub-states of X_0 , as detailed in Sect. 3.6.2. Compare the transitions from X_0 to X_{I_1} in these two cases, and recall Fig. 6 and the weak memoryless property of Poisson process first event times from Sect. 3.1.2

unchanged, but where the different sub-states have functional differences elsewhere in the model. For example, transmission rate reductions from quarantine in SIR type infectious disease models.

One approach to deriving such a model is to condition the dwell time distribution for an intermediate state X_{I_i} on the time already spent in X_0 (as in Feng et al. 2016). We take a slightly different approach and exploit the weak memoryless property of Poisson process 1st event time distributions (see Theorem 1 in Sect. 3.1.2, and the notation used in the previous section) to instead condition the dwell time distribution for intermediate states X_{I_j} on how many of the k_0 events have already occurred when a particle transitions from X_0 to X_{I_j} (rather than conditioning on the exact elapsed time spent in X_0). In this case, since each sub-state of X_0 has iid dwell time distributions that are Poisson process 1st event times with rate $r(t) = \sum_{i=0}^N r_i(t)$, if i of the k_0 events had occurred prior to the transition out of X_0 , then the weak memoryless property of Poisson process 1st event time distributions implies that the remaining time spent in X_{I_j} should follow a $(k_0 - i)$ th event time distribution under an Poisson process with rate $r_0(t)$, thus ensuring that the total time spent in X follows a k_0 th event time distribution with rate $r_0(t)$. With this realization in hand, one can then apply Theorem 6 and Theorem 4 as in the previous section to obtain the desired mean field ODEs, as detailed in the following Theorem, and as illustrated in Fig. 8.

Theorem 9 (Extended LCT with dwell time preserving intermediate states) *Consider the mean field equations for a system of particles entering state X (into sub-state X_0) at rate $\mathcal{I}_X(t)$. As in the previous case, assume the time spent in X_0 follows the minimum of $N + 1$ independent Poisson process k_i th event time distributions with respective rates $r_i(t)$, $i = 0, \dots, N$ (i.e., $T = \min_i(T_i)$). Particles leaving X_0 at time T transition to recipient state Y_ℓ with probability $p_{0\ell}(T)$ if $T = T_0$, or if $T = T_i$ ($i = 1, \dots, N$) to the j th of M_X intermediate sub-states, X_{I_j} , with probability $p_{ij}(T)$. If $T < T_0$, we may define a random variable $K \in \{0, \dots, k_0 - 1\}$ indicating how many events had occurred under the Poisson process associated with T_0 at the time of the transition out of X_0 (at time T). In order to ensure the overall time spent in X follows a Poisson process k_0 th event time distribution with rate $r_0(t)$, it follows that particles entering state, X_{I_j} will remain there for a duration of time that is conditioned on $K = k$ such that the conditional dwell time for that particle in X_{I_j} will be given by a Poisson process $(k_0 - k)$ th event time with rate $r_0(t)$. Finally, assume that particles leaving X via intermediate sub-state X_{I_j} at time t transition to Y_ℓ with probability $q_{j\ell}$, where they remain according to a dwell time determined by survival function $S_\ell(t, \tau)$.*

The corresponding mean field equations are

$$\frac{d}{dt}x_{0(1,\dots,1)}(t) = \mathcal{I}_X(t) - \sum_{i=0}^N r_i(t) x_{0(1,\dots,1)}(t) \tag{66a}$$

$$\frac{d}{dt}x_{0\alpha}(t) = \sum_{i=0}^N r_i(t) x_{0\alpha_{i,-1}}(t) \mathbb{1}_{[a_i>1]}(\alpha) - \sum_{i=0}^N r_i(t) x_{0\alpha}(t) \tag{66b}$$

$$\frac{d}{dt}x_{I_{jk}}(t) = r_0(t) (x_{I_{j,k-1}}(t) \mathbb{1}_{[k>1]} - x_{I_{jk}}(t)) + \sum_{\alpha \in \mathcal{K}_{ij}} r_i(t) x_\alpha(t) p_{ij}(t) \tag{66c}$$

$$\begin{aligned}
 y_\ell(t) = & y_\ell(0)S_\ell(t, 0) + \int_0^t \left(\mathcal{I}_{Y_\ell}(\tau) + \sum_{\alpha \in \mathcal{K}_0} r_0(\tau) x_\alpha(\tau) \right. \\
 & \left. + \sum_{j=1}^{M_X} r_0(\tau) x_{I_{jk_0}}(\tau) q_{j\ell}(\tau) \right) S_\ell(t, \tau) d\tau
 \end{aligned} \tag{66d}$$

where $\mathcal{K} = \{(a_0, a_1, \dots, a_N) \mid a_j \in \{1, \dots, k_j\}\}$, $\alpha \in \mathcal{K} \setminus (1, \dots, 1)$, $j = 1, \dots, M_X$, $k = 1, \dots, k_0$, $\ell = 1, \dots, M_Y$, $\mathcal{K}_i \subset \mathcal{K}$ are the subset of indices where $a_i = k_i$, $\mathcal{K}_{ij} \subset \mathcal{K}_i$ are the subset of indices where $a_i = k_i$ and $a_0 = j$, $x_0(t) = \sum_{\alpha \in \mathcal{K}} x_{0\alpha}(t)$, $x_{iI}(t) = \sum_{j=1}^{k_0} x_{iI_j}(t)$, and $x(t) = x_0(t) + \sum_{i=1}^n x_{iI}(t)$. The $y_\ell(t)$ equations (66d) may be reduced to ODEs, e.g., via Corollary 1.

Proof The proof of Theorem 9 parallels the proof of Theorem 8, but with the following modifications. First, each sub-state of X_{I_j} (for all j) has the same dwell time distribution, namely, they are all 1st event time distributions under a Poisson process with rate $r_0(t)$. Second, upon leaving X_0 where $T = T_i$ and $K(T) = k < k_0$ (i.e., when only $k < k_0$ events have occurred under the 0th Poisson process; see the definition of K in the text above) particles will enter (with probability $p_{ij}(T)$) the j th intermediate state X_{I_j} by entering sub-state $X_{I_{jk}}$ which (due to the weak memorylessness property described in Theorem 1) ensures that, upon leaving X_{I_j} particles will have spent a duration of time that follows the Poisson process k_0 th event time distribution with rate $r_0(t)$. □

Example 7 Consider Example 6 in the previous section, but instead assume that the transition to the intermediate state does not impact the overall time spent in state X , as detailed in Theorem 9. The corresponding mean field ODEs for this case are given by Eq. (67) below [compare Eqs. (67e)–(67g) to Eqs. (65e)–(65h)].

$$\frac{d}{dt} x_{0(1,1)}(t) = \mathcal{I}_X(t) - (r_0 + r_1) x_{0(1,1)}(t) \tag{67a}$$

$$\frac{d}{dt} x_{0(2,1)}(t) = r_0 x_{0(1,1)}(t) - (r_0 + r_1) x_{0(2,1)}(t) \tag{67b}$$

$$\frac{d}{dt} x_{0(1,2)}(t) = r_1 x_{0(1,1)}(t) - (r_0 + r_1) x_{0(1,2)}(t) \tag{67c}$$

$$\frac{d}{dt} x_{0(2,2)}(t) = r_0 x_{0(1,2)}(t) + r_1 x_{0(2,1)}(t) - (r_0 + r_1) x_{0(2,2)}(t) \tag{67d}$$

$$\frac{d}{dt} x_{I_{11}}(t) = r_1 x_{0(1,2)}(t) - r_0 x_{I_{11}}(t) \tag{67e}$$

$$\frac{d}{dt} x_{I_{12}}(t) = r_1 x_{0(2,2)}(t) + r_0 x_{I_{11}}(t) - r_0 x_{I_{12}}(t) \tag{67f}$$

$$\frac{d}{dt} y(t) = \mathcal{I}_Y(t) + r_0 x_{0(2,1)}(t) + r_0 x_{0(2,2)}(t) + r_0 x_{I_{12}}(t) - \mu y(t). \tag{67g}$$

3.7 Generalized Linear Chain Trick (GLCT)

In the preceding sections we have provided various extensions of the Linear Chain Trick (LCT) that describe how the structure of mean field ODE models reflects the assumptions that define corresponding continuous time stochastic state transition models. Each case above can be viewed as a special case of the following more general framework for constructing mean field ODEs, which we refer to as the Generalized Linear Chain Trick (GLCT).

The cases we have addressed thus far share the following stochastic model assumptions, which constitute the major assumptions of the GLCT.

- A1. A focal state (which we call state X) can be partitioned into a finite number of sub-states (e.g., X_1, \dots, X_n), each with independent (across states and particles) dwell time distributions that are either exponentially distributed with rates r_i or, more generally, are distributed as independent 1st event times under nonhomogeneous Poisson processes with rates $r_i(t), i = 1, \dots, n$. Recall the equivalence relation in Sect. 3.5.5.
- A2. Inflow rates into X can be described by non-negative, integrable inflow rates into each of these sub-states (e.g., $\mathcal{I}_{X_1}(t), \dots, \mathcal{I}_{X_n}(t)$), some or all of which may be zero. This includes the case where particles enter X at rate $\Lambda(t)$ and are distributed across sub-states X_i according to the probabilities $\boldsymbol{\rho}(t) = [\rho_1(t), \dots, \rho_n(t)]^T$ (i.e., we let $\mathcal{I}_{X_i}(t) \equiv \rho_i(t) \Lambda(t)$) where $\sum_i \rho_i = 1$.
- A3. Particles that transition out of a sub-state X_i at time t transition either into sub-state X_j with probability $p_{ij}(t)$, or into one of recipient states $Y_\ell, \ell = 1, \dots, m$, with probability $p_{i,n+\ell}$. That is, let $p_{ij}(t)$ denote the probability that a particle leaving state X_i at time t enters either X_j if $j \leq n$ or Y_{j-n} if $j > n$, where $i = 1, \dots, n, j = 1, \dots, n, n + 1, \dots, n + m$.
- A4. Recipient states $Y_\ell, \ell = 1, \dots, m$, also have dwell time distributions defined by survival functions $S_{Y_\ell}(t, \tau)$ and integrable, non-negative inflow rates $\mathcal{I}_{Y_\ell}(t)$ that describe inputs from all other non-X sources.

The GLCT (Theorem 10) describes how to construct mean field ODEs for states X and Y for state transition models satisfying the above assumptions.

Theorem 10 (Generalized Linear Chain Trick) *Consider a stochastic, continuous time state transition model of particles entering state X and transitioning to states $Y_\ell, \ell = 1, \dots, m$, according to the above assumptions A1-A4. Then the corresponding mean field model is given by*

$$\frac{d}{dt} x_i(t) = \mathcal{I}_{X_i}(t) + \sum_{j=1}^n p_{ji}(t) r_j(t) x_j(t) - r_i(t) x_i(t), \quad i = 1, \dots, n, \quad (68a)$$

$$y_\ell(t) = y_\ell(0) S_{Y_\ell}(t, 0) + \int_0^t \left(\mathcal{I}_{Y_\ell}(\tau) + \sum_{j=1}^n r_j(\tau) x_j(\tau) p_{j,n+\ell}(\tau) \right) S_{Y_\ell}(t, \tau) d\tau \quad (68b)$$

where $x(t) = \sum_{i=1}^n x_i(t)$, and we assume non-negative initial conditions $x_i(0) = x_{i0}$, $y_\ell(0) = y_{\ell 0}$. Note that the $y_\ell(t)$ equations might be reducible to ODEs, e.g., via Corollary 1 or other results presented above.

Furthermore, Eq. (68a) may be written in vector form where $\mathbf{P}_X(t) = (p_{ij}(t))$ ($i, j \in \{1, \dots, n\}$) is the $n \times n$ matrix of (potentially time-varying) probabilities describing which transitions out of X_i at time t go to X_j (likewise, one can define $\mathbf{P}_Y(t) = (p_{ij}(t))$, $i \in \{1, \dots, n\}$, $j \in \{n + 1, \dots, n + m\}$, which is the $n \times m$ matrix of probabilities describing which transitions from X_i at time t go to Y_{j-n}), $\mathcal{I}_X(t) = [\mathcal{I}_{X_1}, \dots, \mathcal{I}_{X_n}]^T$, $\mathbf{R}(t) = [r_1(t), \dots, r_n(t)]^T$, and $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$ which yields

$$\frac{d}{dt} \mathbf{x}(t) = \mathcal{I}_X(t) + \mathbf{P}_X(t)^T (\mathbf{R}(t) \circ \mathbf{x}(t)) - \mathbf{R}(t) \circ \mathbf{x}(t). \tag{69}$$

where \circ indicates the Hadamard (element-wise) product.

Proof The proof of the theorem above follows directly from applying Theorem 4 to each sub-state. □

Example 8 (Dwell time given by the maximum of independent Erlang random variables) We here illustrate how the GLCT can provide a conceptually simpler framework for deriving ODEs relative to derivation from mean field integral equations by assuming the X dwell time obeys the maximum of multiple Erlang distributions. While the survival function for this distribution is not straightforward to write down, it is fairly straightforward to construct a Markov Chain that yields such a dwell time distribution (see Fig. 9).

Recall that, in Sect. 3.5.4, we considered a dwell time given by the minimum of N Erlang distributions. Here we instead consider the case where the dwell time distribution is given by the maximum of multiple Erlang distributions, $T = \max(T_1, T_2)$ where $T_i \sim \text{Erlang}(r_i, 2)$. For simplicity, assume the dwell time in a single recipient state Y is exponential with rate μ . We again partition X according to which events (under the two independent homogeneous Poisson processes associated with each of T_1 and T_2) particles are awaiting, and index those sub-states accordingly (see Fig. 9). These sub-states are $X_{11}, X_{21}, X_{12}, X_{*1}, X_{22}, X_{1*}, X_{*2}$, and X_{2*} , where a ‘*’ in the i th index position indicates that particles in that sub-state have already had the i th Poisson process reach the k_i th event (in this case, the 2nd event). Each such sub-state has exponentially distributed dwell times, but rates for these dwell time distributions differ (unlike the cases in Sect. 3.5.4 where all sub-states had the same rate): the Poisson process rates for sub-states X_{11}, X_{21}, X_{12} , and X_{22} are $r = r_1 + r_2$ (see Fig. 9 and compare to Theorem 6 and Fig. 6), but the rate for the states X_{1*} and X_{2*} (striped circles in Fig. 9) are r_1 , and for X_{*1} and X_{*2} (shaded circles in Fig. 9) are r_2 .

In the context of the GLCT, let $\mathbf{x}(t) = [x_{11}(t), x_{21}(t), x_{12}(t), x_{*1}(t), x_{22}(t), x_{1*}(t), x_{*2}(t), x_{2*}(t)]^T$ then by the assumptions above $\mathbf{R}(t) = [r, r, r, r_2, r, r_1, r_2, r_1]^T$, $\rho(t) = [1, 0, \dots, 0]^T$ and hence $\mathcal{I}_X(t) = [\mathcal{I}_X(t), 0, \dots, 0]^T$. Denote $p_1 \equiv r_1/r$ and $p_2 \equiv r_2/r$ (à la Theorem 7 in Sect. 3.5.5). Then the first eight rows of 9×9 matrix \mathbf{P} are given by

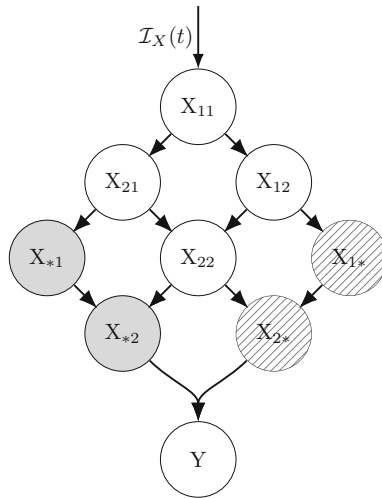


Fig. 9 The X sub-state structure for Example 8 where X dwell time distribution follows the maximum of two Erlang random variables with rates r_1 and r_2 , respectively, and shape parameters $k_i = 2$. As in Sect. 3.5.4, where the minimum is assumed instead of the maximum, X can be partitioned using indices based on organizing particles by which events they are awaiting under each Poisson process associated with each Erlang distribution. Upon reaching the target event (here, the 2nd event) under any given Poisson process, particles transition to sub-states with an asterisk in the corresponding index position (e.g., see figure). In general, these sub-states all have a dwell times given by the 1st event time under a Poisson process, but with differing rates (see Example 8): here they follow exponential distributions with either rate $r = r_1 + r_2$ (white backgrounds), rate r_2 (gray backgrounds), or rate r_1 (lined backgrounds)

$$[\mathbf{P}_X \mid \mathbf{P}_Y] = \begin{bmatrix} 0 & p_1 & p_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_1 & p_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_1 & p_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_1 & p_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{70}$$

Thus, by the GLCT (Theorem 10), the corresponding mean field ODEs are

$$\frac{d}{dt}x_{11}(t) = \mathcal{I}_X(t) - r x_{11}(t) \tag{71a}$$

$$\frac{d}{dt}x_{21}(t) = r_1 x_{11}(t) - r x_{21}(t) \tag{71b}$$

$$\frac{d}{dt}x_{12}(t) = r_2 x_{11}(t) - r x_{12}(t) \tag{71c}$$

$$\frac{d}{dt}x_{*1}(t) = r_1 x_{21}(t) - r_2 x_{*1}(t) \tag{71d}$$

$$\frac{d}{dt}x_{22}(t) = r_2 x_{21}(t) + r_1 x_{12}(t) - r x_{22}(t) \tag{71e}$$

$$\frac{d}{dt}x_{1*}(t) = r_2 x_{12}(t) - r_1 x_{1*}(t) \quad (71f)$$

$$\frac{d}{dt}x_{*2}(t) = r_2 x_{*1}(t) + r_1 x_{22}(t) - r_2 x_{*2}(t) \quad (71g)$$

$$\frac{d}{dt}x_{2*}(t) = r_1 x_{1*}(t) + r_2 x_{22}(t) - r_1 x_{2*}(t) \quad (71h)$$

$$\frac{d}{dt}y(t) = r_1 x_{2*}(t) + r_2 x_{*2}(t) - \mu y(t). \quad (71i)$$

3.7.1 GLCT for phase-type distributions

The GLCT above extends the LCT to a very flexible family of dwell time distributions known as (continuous) *phase-type distributions* (Asmussen et al. 1996; Pérez and Riaño 2006; Osogami and Harchol-Balter 2006; Thummler et al. 2006; Reinecke et al. 2012a; Horváth et al. 2012; Komárková 2012; Horváth et al. 2016; Okamura and Dohi 2015; Horváth and Telek 2017), i.e., the hitting time distributions for Continuous Time Markov Chains (CTMC). These CTMC hitting time distributions include the hypoexponential distributions, hyper-exponential and hyper-Erlang distributions, and generalized Coxian distributions (Reinecke et al. 2012a; Horváth et al. 2016). Importantly, these distributions can be fit to data or can be used to approximate other named distributions (e.g., see Horváth and Telek 2017; Osogami and Harchol-Balter 2006; Altioik 1985; Pérez and Riaño 2006; Komárková 2012; Reinecke et al. 2012b, and related publications). As detailed below, this enables modelers to incorporate a much broader set of dwell time distributions into ODEs than is afforded by the standard LCT.

Consider the assumptions of Theorem 10 above. Assume vectors $\rho(t) = \rho$ and $\mathbf{R}(t) = \mathbf{R}$, and matrices $\mathbf{P}_X(t) = \mathbf{P}_X$, and $\mathbf{P}_Y(t) = \mathbf{P}_Y$ are all constant. As above, assume the probability of entering states in Y is zero, thus our initial distribution vector for this CTMC (with $n + m$ states) is fully determined by the (length n) vector ρ . Also assume—just to define the CTMC that describes transitions among transient states X up to (but not after) entering states Y—that each state in Y is absorbing (i.e., $p_{ii} = 1$, $i > n$). Then the X dwell time distribution follows the hitting time distribution for a CTMC with initial distribution vector ρ and $(n + m) \times (n + m)$ transition probability matrix

$$\mathbf{P} = \begin{bmatrix} \mathbf{P}_X & \mathbf{P}_Y \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \quad (72)$$

To clearly state the GLCT for phase-type distributions, we must reparameterize the above CTMC. First, there is an equivalent parameterization of this CTMC which corresponds to thinning the underlying Poisson processes so that we only track transitions between distinct states, and ignore when an individual leaves and instantly returns to its current state (this thinned process is often called the embedded jump process).

The rate for the thinned process that determines a particle's dwell time in transient state i goes from r_i to

$$\lambda_i \equiv r_i (1 - p_{ii}) \tag{73}$$

If $0 \leq p_{ii} < 1$, the transition probabilities out of state i then get normalized to

$$\tilde{p}_{ij} \equiv \frac{p_{ij}}{(1 - p_{ii})}, \quad \text{for } j \neq i, \text{ and } \tilde{p}_{ii} = 0. \tag{74}$$

The rows for absorbing states (Y) remain unchanged, i.e., $\tilde{p}_{ij} = p_{ij}$ for $i > n$. The resulting transition probability matrix $\tilde{\mathbf{P}}$ and rate vector λ define the embedded jump process description of the CTMC with transition probability matrix \mathbf{P} and rate vector \mathbf{R} (initial probability vector ρ is the same for both representations of this CTMC).

Lastly, this CTMC can again be reparameterized by combining the jump process transition probability matrix $\tilde{\mathbf{P}}$ and rate vector λ to yield this CTMC's *transition rate matrix* (also sometimes called the *infinitesimal generator matrix* or simply the *generator matrix*) which is defined as follows. Let matrix \mathbf{G} be the same dimension as \mathbf{P} (and thus, $\tilde{\mathbf{P}}$) and let the first n terms in the diagonal of \mathbf{G} be the negative of the jump process rates, $-\lambda$ (i.e., $G_{ii} = -\lambda_i, i \leq n$). Let the off diagonal entries of the first n rows of \mathbf{G} be the jump process transition probabilities \tilde{p}_{ij} multiplied by the i th rate λ_i (i.e., $G_{ij} = \lambda_i \tilde{p}_{ij}, j \neq i$). Thus, the first row of \mathbf{G} is

$$[-\lambda_1 \quad \tilde{p}_{12}\lambda_1 \quad \cdots \quad \tilde{p}_{1n}\lambda_1 \quad \cdots \quad \tilde{p}_{1,n+m}\lambda_1]$$

and so on. Since the transition rates out of absorbing states (e.g., the last m rows of \mathbf{G}) are 0, \mathbf{G} has the form

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_X & \mathbf{G}_Y \\ 0 & 0 \end{bmatrix}. \tag{75}$$

Note that $\tilde{\mathbf{P}}$ and λ can be recovered from \mathbf{G} using the definitions above.

This third parameterization of the given CTMC, determined solely by initial distribution ρ and transition rate matrix \mathbf{G} , can now be used to formally describe the phase-type distribution associated with this CTMC, i.e., the distribution of time spent in the transient states X before hitting absorbing states Y. Specifically, the phase-type distribution density function and CDF are

$$f(t) = \rho^T e^{t \mathbf{G}_X} (-\mathbf{G}_X \mathbf{1}) \tag{76a}$$

$$F(t) = 1 - \rho^T e^{t \mathbf{G}_X} \mathbf{1} \tag{76b}$$

where $\mathbf{1}$ is a $n \times 1$ vector of ones. Importantly, this distribution depends only on the $n \times n$ matrix \mathbf{G}_X and length n initial distribution vector ρ .

The GLCT for phase-type distributions can now be stated as follows.

Corollary 2 (GLCT for phase-type distributions) *Assume particles enter state X at rate $\Lambda(t)$ and that the dwell time distribution for a state X follows a continuous phase-type distribution given by the $n \times 1$ initial probability vector ρ and $n \times n$ matrix \mathbf{G}_X .*

Let $\mathcal{I}_{X_i}(t) = \rho_i \Lambda(t)$ and $\mathcal{I}_{\mathbf{X}}(t) = [\mathcal{I}_{X_1}, \dots, \mathcal{I}_{X_n}]^T$. Then Eq. (69) in Theorem 10 becomes

$$\frac{d}{dt} \mathbf{x}(t) = \mathcal{I}_{\mathbf{X}}(t) + \mathbf{G}_{\mathbf{X}}^T \mathbf{x}(t). \tag{77}$$

Example 9 (Serial LCT and hypoexponential distributions) Assume the dwell time in state \mathbf{X} is given by the sum of independent (not identically distributed) Erlang distributions or, more generally, Poisson process k_i th event time distributions with rates $r_i(t)$, i.e., $T = \sum_i T_i, i = 1, \dots, N$ (note the special case where all $k_i = 1$ and $r_i(t) = r_i$ are constant, which yields that T follows a hypoexponential distribution). Let $n = \sum_i k_i$ and further assume particles go to Y_ℓ with probability p_ℓ upon leaving $\mathbf{X}, \ell = 1, \dots, m$. Using the GLCT framework above, this corresponds to partitioning \mathbf{X} into sub-states \mathbf{X}_j , where $j = 1, \dots, n$, and

$$\mathbf{R}(t) = [r_1(t), r_1(t), \dots, r_2(t), \dots, r_n(t)]^T \tag{78}$$

where the first k_1 elements of $\mathbf{R}(t)$ are $r_1(t)$, the next k_2 are $r_2(t)$, etc., and

$$\mathbf{P}_X = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}_{n \times n}, \quad \mathbf{P}_Y = \begin{bmatrix} 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ p_1 & p_2 & \dots & p_m \end{bmatrix}_{n \times \ell}. \tag{79}$$

By the GLCT (Theorem 10), using $r_{(j)}(t)$ to denote the j th element of $\mathbf{R}(t)$ in Eq. (78), the corresponding mean field equations are

$$\frac{d}{dt} x_1(t) = \mathcal{I}_X(t) - r_1(t) x_1(t) \tag{80a}$$

$$\frac{d}{dt} x_j(t) = r_{(j-1)}(t) x_{j-1}(t) - r_{(j)}(t) x_j(t), \text{ for } j \geq 1, \tag{80b}$$

$$y_\ell(t) = y_\ell(0) S_{Y_\ell}(t, 0) + \int_0^t \left(\mathcal{I}_{Y_\ell}(\tau) + \sum_{j=1}^m r_{(j)}(\tau) x_j(\tau) p_j \right) S_{Y_\ell}(t, \tau) d\tau. \tag{80c}$$

Note, the phase-type distribution form of the hypoexponential distribution could be used with Corollary 2 to derive the x_i equations in Eq. (80).

Example 10 (SIR model with Phase-Type Duration of Infectiousness.) Consider the simple SIR model given by Eq. (2) with mass action transmission $\lambda(t) = \beta I(t)$. Suppose the assumed exponentially distributed dwell times in state I were to be

replaced by a phase-type distribution with initial distribution vector α and matrix \mathbf{A} (note that, in some cases, it is possible to match the first three or more moments using only a 2×2 or 3×3 matrix \mathbf{A} and note that Matlab and Python routines for making such estimates are freely available in Horváth and Telek 2017). Then, letting \mathbf{x} be the vector of substates of I and $I(t) = x_1(t) + \dots + x_n(t)$, by the GLCT for phase-type distributions (Corollary 2) the corresponding mean field ODE model, with $R(t) = N_0 - S(t) - I(t)$, is

$$\frac{d}{dt} S(t) = -\beta I(t) S(t) \quad (81a)$$

$$\frac{d}{dt} \mathbf{x}(t) = \alpha \beta I(t) S(t) + \mathbf{A}^T \mathbf{x}(t). \quad (81b)$$

4 Discussion

ODEs are arguably the most popular framework for modeling continuous time dynamical systems in applications. This is in part due to the relative ease of building, simulating, and analyzing ODE models, which makes them very accessible to a broad range of scientists and mathematicians. The above results generalize the Linear Chain Trick (LCT), and detail how to construct mean field ODE models for a broad range of scenarios found in applications based on explicit individual-level stochastic model assumptions. Our hope is that these contributions improve the speed and efficiency of constructing mean field ODE models, increase the flexibility to make more appropriate dwell time assumptions, and help clarify (for both modelers and those reading the results of their work) how individual-level stochastic assumptions are reflected in the structure of mean field ODE model equations. We have provided multiple novel theorems that describe how to construct such ODEs directly from underlying stochastic model assumptions, without formally deriving them from an explicit stochastic model or from intermediate integral equations. The Erlang distribution recursion relation (Lemma 1) that drives the LCT has been generalized to include the time-varying analogues of Erlang distributions—i.e., k th event time distributions under nonhomogeneous Poisson processes (Lemma 2)—and distributions that reflect “competing Poisson process event times” defined as the minimum of a finite number of independent Poisson process event times (Lemma 3). These new lemmas, and our generalization of the memorylessness property of the exponential distribution (which we refer to in Sect. 3.1.2 as the *weak memorylessness* property of nonhomogeneous Poisson process 1st event time distributions) together allow more flexible dwell time assumptions to be incorporated into mean field ODE models. We have also introduced a novel Generalized Linear Chain Trick (GLCT; Theorem 10 in Sect. 3.7) which complements previous extensions of the LCT (e.g., Jacquez and Simon 2002; Diekmann et al. 2017) and allows modelers to construct mean field ODE models for a broad range of dwell time assumptions and associated sub-state configurations (e.g., conditional dwell time distributions for intermediate sub-state transitions), including the phase-type family of distributions and their time-varying analogues as detailed in Sect. 3.7.1. The GLCT also provides a framework for considering other scenarios not specifically addressed by

the preceding results, as illustrated by Example 8 in which the dwell time distribution follows the maximum of multiple Erlang distributions.

These results not only provide a framework to incorporate more accurate dwell time distributions into ODE models, but also hopefully encourage more comparative studies, such as Feng et al. (2016), that explore the dynamic and application-specific consequences of incorporating non-Erlang dwell time distributions, and conditional dwell time distributions, into ODE models. More study is needed, for example, comparing integral equation models with non-exponential dwell time distributions and their ODE counterparts based on approximating those distributions with Erlang and phase-type distributions.

The GLCT also permits modelers to incorporate the flexible phase-type distributions (i.e., the hitting-time distributions for Continuous Time Markov Chains) into ODE models. This family of probability distributions includes mixtures of Erlang distributions (a.k.a. hyper-Erlang distributions), the minimum or maximum of multiple Erlang distributions, the hypoexponential distributions, generalized Coxian distributions, and others (Reinecke et al. 2012a; Horváth et al. 2016). While the phase-type distributions are currently mostly unknown to mathematical biologists, they have received some attention in other fields and modelers can take advantage of existing methods that have been developed to fit phase-type distributions to other distributions on \mathbb{R}^+ and to data (see Sect. 3.7.1 for references). These results provide a flexible framework for approximating dwell time distributions, and incorporating those empirically or analytically derived dwell time distributions into ODE models.

There are some additional considerations, and potential challenges to implementing these results in applications, that are worth addressing. First, the increase in the number of state variables may lead to both computational and analytical challenges, however we have a growing number of tools at our disposal for tackling higher dimensional systems (e.g., van den Driessche and Watmough 2002; Guo et al. 2008; Li and Shuai 2010; Du and Li 2014; Bindel et al. 2014). We intend to explore the costs associated with this increase in dimensionality in future publications. Second, it is tempting to assume that the sub-states resulting from the above theorems correspond to some sort of sub-state structure in the actual system being modeled. This is not necessarily the case, and we should be cautious about interpreting these sub-states as evidence of, e.g., cryptic population structure. Third, some of the above theorems make a simplifying assumption that, either at time $t = 0$ or upon entry into X , the initial distribution of particles is only into the first sub-state. This may not be the appropriate assumption to make in some applications, but it is fairly straight forward to modify these these initial condition assumptions within the context of the GLCT. Fourth, it may be appropriate in some applications to avoid mean field models, and instead analyze the stochastic model dynamics directly [e.g., see Allen (2010, 2017) and references therein]. Lastly, the history of successful attempts to garner scientific insights from mean field ODE models seems to suggest that such distributional refinements are unnecessary. However, this is clearly not always the case, as evidenced by the many instances in which modelers have abandoned ODEs and instead opted to use integral equations to model systems with non-Erlang dwell time distributions. We hope these results will facilitate more rigorous comparisons between such models and their simplified and/or ODE counterparts, like

Feng et al. (2016) and Piotrowska and Bodnar (2018), to help clarify when the use of these ODE models is warranted.

In closing, these novel extensions of the LCT help clarify how underlying assumptions are reflected in the structure of mean field ODE models, and provide a means for incorporating more flexible dwell time assumptions into mean field ODE models directly from first principles without a need to derive ODEs from stochastic models or intermediate mean field integral equations. The Generalized Linear Chain Trick (GLCT) provides both a conceptual framework for understanding how individual-level stochastic assumptions are reflected in the structure of mean field model equations, and a practical framework for incorporating exact, approximate, or empirically derived dwell time distributions and related assumptions into mean field ODE models.

Acknowledgements The authors thank Michael H. Cortez, Jim Cushing, Marisa Eisenberg, Jace Gilbert, Zoe Haskell, Tomasz Kozubowski, Miles Moran, Catalina Medina, Amy Robards, Deena R. Schmidt, Joe Tien, Narae Wadsworth, and an anonymous referee for comments and suggestions that improved this work.

Funding This work was supported by funds provided to PJH by the University of Nevada, Reno (UNR) Office of Research and Innovation, and by a grant from the Sloan Scholars Mentoring Network of the Social Science Research Council with funds provided by the Alfred P. Sloan Foundation.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided 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.

Appendix A Deterministic models as mean field equations

We here give a brief description of the process of deriving deterministic mean field equations from stochastic first principles.

Consider a stochastic model of discrete particles (e.g., individual organisms in a population, molecules in a solution, etc.) that transition among a finite number of n states in continuous time. Let the state variables of interest be the amount of particles in each state. This set of counts in each state can be thought of as a *state vector* in the *state space* $\mathbb{N}^n \subset \mathbb{R}^n$, and our model describes the (stochastic) rules governing transitions from one state vector to the next (i.e., from a given state there is some probability distribution across the state space describing how the system will proceed).

Mean field models essentially average that distribution, and thus describe the mean state transitions from any given state of the system. That is, for a given state in \mathbb{R}^n , we can think of a probability distribution that describes where the system would move from that point in \mathbb{R}^n , and find the mean transition direction in state space, which then defines a deterministic dynamical system on \mathbb{R}^n which we refer to as a mean field model for the given stochastic process. To derive a continuous time mean field model, we do this for an arbitrary discrete step size Δt so we can then take the limit as $\Delta t \rightarrow 0$.

A.1 Deriving mean field integral equations and ODEs

Here we illustrate one approach to derive mean field equations starting from an explicit stochastic model using a simple toy example. To do this, we begin by describing a discrete time approximation (with arbitrarily small time step Δt) of a stochastic system of particles transitioning among multiple states, then derive the corresponding discrete-time mean field model which then yields the desired mean field model by taking the limit as $\Delta t \rightarrow 0$. Other more systematic approaches exist, e.g., see Kurtz (1970, 1971).

Consider a large number of particles that begin (at time $t = 0$) in state W and then each transitions to state X after an exponentially distributed amount of time (with rate a). Particles remain in state X according to a Erlang(r, k) distributed delay before entering state Y. They then go to state Z after an exponentially distributed time delay with rate μ . Let $\tilde{w}(t), \tilde{x}(t), \tilde{y}(t),$ and $\tilde{z}(t)$ be the amount in each of the corresponding states at time $t \geq 0$, with $w(0) = N_0$, and $x(0) = y(0) = z(0) = 0$. Then, as we will derive below, the corresponding continuous time mean field model is given by Eq. (A1) where the state variables $w, x, y,$ and z correspond to the amount in each of the corresponding states, and we assume the initial conditions $w(0) = w_0 \gg 1$ and $x(0) = y(0) = z(0) = 0$.

$$\frac{d}{dt}w(t) = -a w(t) \tag{A1a}$$

$$x(t) = \int_0^t \underbrace{a w(s)}_{\mathcal{I}(s)} \underbrace{S_r^k(t-s)}_{\text{proportion remaining}} ds \tag{A1b}$$

$$y(t) = \int_0^t \underbrace{\left(\int_0^\tau a w(s) g_r^k(\tau-s) ds \right)}_{\text{Net input rate at time } \tau} S_\mu^1(t-\tau) d\tau \tag{A1c}$$

$$\frac{d}{dt}z(t) = \mu y(t) \tag{A1d}$$

First, to derive the linear ODEs (A1a) and (A1d), note that the number of particles that transition from state W to state X in a short time interval $(t, t + \Delta t)$ is binomially distributed: if we think of a transition from W to X as a “success” then the number of “trials” $n = \tilde{w}(t)$ and the probability of success p is given by the exponential CDF value $p = 1 - \exp(-a\Delta t)$. For sufficiently small Δt this implies $p = a\Delta t + O(\Delta t^2)$. Let $w(t + s) = E(\tilde{w}(t + s) | \tilde{w}(t))$ for $s \geq 0$. Since the expected value of a binomial random variable is np it follows that

$$\begin{aligned} w(t + \Delta t) - w(t) &\equiv E(\tilde{w}(t + \Delta t) - \tilde{w}(t) | \tilde{w}(t)) \\ &= -a w(t) \Delta t + O(\Delta t^2). \end{aligned} \tag{A2}$$

Dividing both sides of (A2) by Δt and then taking the limit as $\Delta t \rightarrow 0$ yields

$$\frac{d}{dt}w(t) = -a w(t). \tag{A3}$$

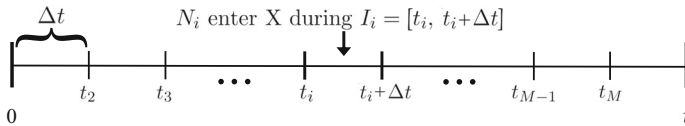


Fig. 10 Illustration of discretizing the time interval $[0, t]$ into M bins (each $\Delta t = t/M$ wide) in order to derive the integral equations (A1b) and (A1c)

Similarly, define $z(t)$ in terms of $\tilde{z}(t)$ then it follows that

$$\frac{d}{dt}z(t) = \mu y(t). \tag{A4}$$

Next, we derive the integral equations (A1b) and (A1c) by similarly deriving a discrete time mean field model and then taking its limit as bin width $\Delta t \rightarrow 0$ (i.e., as the number of bins $M \rightarrow \infty$).

Partition the interval $[0, t]$ into $M \gg 1$ equally wide intervals of width $\Delta t \equiv t/M$ (see Fig. 10). Let $I_i = ((i - 1)\Delta t, i\Delta t]$ denote the i th such time interval ($i = 1, \dots, M$) and let $t_i = (i - 1)\Delta t$ denote the start time of the i th interval. We may now account for the number transitioning into and out of state X during time interval I_i , and sum these values to compute $x(t)$.

The number in state X at time t ($\tilde{x}(t)$) is the number that entered state X between time 0 and t , less the number that transitioned out of X before time t (for now, assume $x(0) = 0$). A particle that enters state X at time $s \in (0, t)$ will still be in state X according to a Bernoulli random variable with $p = S_r^k(t - s)$ (the expected proportion under the given gamma distribution). Therefore, to compute $\tilde{x}(t)$ we can sum over our M intervals and add up the number that entered state X during interval I_i and, from each of those M cohorts, count how many remain in X at time t . Specifically, the number entering X during the i th interval $[t_i, t_i + \Delta t]$ is given by $N_i \equiv \tilde{w}(t_i + \Delta t) - \tilde{w}(t_i)$ (see Fig. 10), and thus the number remaining in X at time t is the sum of the number remaining at time t from each such cohort (i.e., the sum over $i = 1$ to M) where the number remaining in X at t from each cohort follows a compound binomial distribution given by the sum of N_i Bernoulli random variables B_X each with probability $p = S_r^k(t - t_i) + O(\Delta t)$. This defines our stochastic state transition model, which yields a mean field model as follows.

The expected amount entering X during $[t_i, t_i + \Delta t]$ is $E(N_i) = E(\tilde{w}(t_i) - \tilde{w}(t_i + \Delta t)) = a w(t_i)\Delta t$, and the expected proportion of the i th cohort remaining at time t is $E(B_X) = S_r^k(t - t_i) + O(\Delta t)$. Thus, the expected number from the i th cohort remaining in X at time t is $E(N_i)E(B_X) = a w(t_i) S_r^k(t - t_i) \Delta t + O(\Delta t^2)$. Summing these expected values over all intervals, and letting $\Delta t \rightarrow 0$, yields

$$\begin{aligned} x(t) = E(\tilde{x}(t)) &= \lim_{M \rightarrow \infty} \sum_{i=1}^M a w(t_i) S_r^k(t - t_i) \Delta t + O(\Delta t^2) \\ &= \int_0^t a w(s) S_r^k(t - s) ds. \end{aligned} \tag{A5}$$

To calculate $y(t)$, let K_j be the number entering state Y during interval I_j that are still in state Y at time t . As above, K_j can be calculated by summing (over I_1 to I_{j-1}) the number in each cohort that entered state X during I_i then transitioned to state Y during time interval I_j and are still in state Y at t . Therefore K_j can be written as the sum of $j - 1$ compound distributions given by counting how many of the N_i particles that entered state X during I_i then transitioned to state Y during interval I_j and then persisted until time t without transitioning to Z. To count these, notice that each such particle entering state X during I_i , state Y during I_j and persisting in Y at time t follows a Bernoulli random variable B_{ij} with probability

$$p_{ij} = \underbrace{(S_r^k(t_j - t_i) - S_r^k(t_{j+1} - t_i) + O(\Delta t))}_{P(X \rightarrow Y \text{ in } I_j | W \rightarrow X \text{ in } I_i)} \underbrace{(S_\mu^1(t - t_j) + O(\Delta t))}_{P(\text{still in Y at } t)}. \quad (A6)$$

Therefore, the number of particles that entered Y at I_j and remain in state Y at time t , K_j , can be written as a compound random variable $K_j = \sum_{i=1}^{j-1} \sum_{k=1}^{N_i} B(p_{ij})$. The expected K_j is thus

$$\begin{aligned} E(K_j) &= \sum_{i=1}^{j-1} E(N_i)E(B_{ij}) \\ &= \sum_{i=1}^{j-1} (a w(t_i) \Delta t) [(S_r^k(t_j - t_i) - S_r^k(t_{j+1} - t_i)) S_\mu^1(t - t_j) + O(\Delta t)] \\ &= \sum_{i=1}^{j-1} a w(t_i) \frac{(S_r^k(t_j - t_i) - S_r^k(t_j - t_i + \Delta t))}{\Delta t} S_\mu^1(t - t_j) \Delta t^2 + O(\Delta t^2) \\ &= \sum_{i=1}^{j-1} a w(t_i) \frac{(G_r^k(t_j - t_i + \Delta t) - G_r^k(t_j - t_i))}{\Delta t} S_\mu^1(t - t_j) \Delta t^2 + O(\Delta t^2). \end{aligned} \quad (A7)$$

Summing over all intervals and letting $\Delta t \rightarrow 0$ ($M \rightarrow \infty$) gives

$$\begin{aligned} y(t) &= \lim_{M \rightarrow \infty} E(\tilde{y}(t)) = \lim_{M \rightarrow \infty} E\left(\sum_{j=1}^M K_j\right) = \lim_{M \rightarrow \infty} \sum_{j=1}^M E(K_j) \\ &= \lim_{M \rightarrow \infty} \sum_{j=1}^M \sum_{i=1}^{j-1} a w(t_i) \frac{(G_r^k(t_j - t_i + \Delta t) - G_r^k(t_j - t_i))}{\Delta t} \Delta t \\ &\quad \cdot S_\mu^1(t - t_j) \Delta t + O(\Delta t^2) \\ &= \int_0^t \left(\int_0^\tau a w(s) g_r^k(\tau - s) ds\right) S_\mu^1(t - \tau) d\tau. \end{aligned} \quad (A8)$$

Applying Theorem 2, the mean field ODEs equivalent to Eq. (A1), where $x(t) = \sum_{j=1}^k x_j(t)$, are given by

$$\frac{d}{dt}w(t) = -a w(t) \tag{A9a}$$

$$\frac{d}{dt}x_1(t) = a w(t) - r x_1(t) \tag{A9b}$$

$$\frac{d}{dt}x_j(t) = r x_{j-1}(t) - r x_j(t), \quad \text{for } j = 2, \dots, k \tag{A9c}$$

$$\frac{d}{dt}y(t) = r x_k(t) - \mu y(t) \tag{A9d}$$

$$\frac{d}{dt}z(t) = \mu y(t). \tag{A9e}$$

Appendix B Erlang mixture approximation of Gamma(α, β)

Here we give a simple example of analytically approximating a gamma distribution with a mixture of two Erlang distributions by moment matching.

Suppose random variable T follows a gamma(α, β) distribution, which has mean $\mu = \beta/\alpha$ and variance $\sigma^2 = \mu/\alpha$, and shape β is not an integer. One can approximate this gamma distribution with a mixture of two Erlang distributions that yields the same mean and variance.

These Erlang distributions are $T_\downarrow \sim \text{Erlang}(r_\downarrow, k_\downarrow)$ and $T_\uparrow \sim \text{Erlang}(r_\uparrow, k_\uparrow)$ where the shape parameters are obtained by rounding β down and up, respectively, to the nearest integer ($k_\downarrow \equiv \lfloor \beta \rfloor$ and $k_\uparrow \equiv \lceil \beta \rceil$) and the rate parameters are given by $r_\downarrow = \alpha \frac{\lfloor \beta \rfloor}{\beta}$ ($r_\downarrow = \frac{k_\downarrow}{\mu}$) and $r_\uparrow = \alpha \frac{\lceil \beta \rceil}{\beta}$ ($r_\uparrow = \frac{k_\uparrow}{\mu}$). This ensures that T_\downarrow and T_\uparrow have mean μ .

To calculate their variance, let $p \equiv \lceil \beta \rceil - \beta$ and $q \equiv \beta - \lfloor \beta \rfloor$ (note $p + q = 1$). By rounding shape β down(up) the resulting Erlang distribution has higher(lower) variance, i.e.,

$$\sigma_\downarrow^2 \equiv \frac{\mu}{r_\downarrow} = \sigma^2 \left(1 + \frac{q}{k_\downarrow} \right) \quad \text{and} \quad \sigma_\uparrow^2 \equiv \frac{\mu}{r_\uparrow} = \sigma^2 \left(1 - \frac{p}{k_\uparrow} \right). \tag{B1}$$

To calculate the mixing proportion, let the mixture distribution $T_\rho = B_\rho T_\downarrow + (1 - B_\rho) T_\uparrow$, where B_ρ is a Bernoulli random variable with $P(B_\rho = 1) = \rho$ and

$$\rho = \frac{p/k_\uparrow}{p/k_\uparrow + q/k_\downarrow}. \tag{B2}$$

This Erlang mixture has the desired mean $E(T_\rho) = \rho \mu + (1 - \rho) \mu = \mu$ and variance σ^2 , since

$$\begin{aligned}
\text{Var}(T_\rho) &= E((B_\rho T_\downarrow + (1 - B_\rho) T_\uparrow)^2) - \mu^2 \\
&= E((B_\rho T_\downarrow)^2) + E(((1 - B_\rho) T_\uparrow)^2) - \mu^2 \\
&= E(B_\rho^2)E(T_\downarrow^2) + E((1 - B_\rho)^2)E(T_\uparrow^2) - \mu^2 \\
&= E(B_\rho)E(T_\downarrow^2) + E(1 - B_\rho)E(T_\uparrow^2) - \mu^2 \\
&= \rho E(T_\downarrow^2) + (1 - \rho) E(T_\uparrow^2) - \mu^2 \\
&= \rho \sigma_\downarrow^2 + (1 - \rho) \sigma_\uparrow^2 \tag{B3} \\
&= \rho \sigma^2 \left(1 + \frac{q}{k_\downarrow}\right) + (1 - \rho) \sigma^2 \left(1 - \frac{p}{k_\uparrow}\right) \\
&= \sigma^2 + \sigma^2 \left(\frac{\rho q}{k_\downarrow} - \frac{(1 - \rho) p}{k_\uparrow}\right) \\
&= \sigma^2 + \sigma^2 \left(\frac{p/k_\uparrow \cdot q/k_\downarrow}{p/k_\uparrow + q/k_\downarrow} - \frac{q/k_\downarrow \cdot p/k_\uparrow}{p/k_\uparrow + q/k_\downarrow}\right) = \sigma^2.
\end{aligned}$$

Numerical comparisons suggest this mixture is a very good approximation of the target $\text{gamma}(\alpha, \beta)$ distribution for shape β values larger than roughly 3 to 5, depending (to a lesser extent) on α .

Alternatively, to approximate a $\text{gamma}(\alpha, \beta)$ distribution with a mixture of Erlang distributions as described above, one could also select the mixing probabilities by, for example, using an alternative metric such as a distance in probability space (e.g., see Rachev 1991), e.g., the L^∞ -norm on their CDFs, or information-theoretic quantities such as KL or Jensen-Shannon divergence.

References

- Allen LJS (2010) An introduction to Stochastic processes with applications to biology, 2nd edn. Chapman and Hall/CRC, Boca Raton
- Allen LJ (2017) A primer on stochastic epidemic models: formulation, numerical simulation, and analysis. *Infect Dis Model* 2(2):128–142. <https://doi.org/10.1016/j.idm.2017.03.001>
- Altiock T (1985) On the phase-type approximations of general distributions. *IEE Trans* 17(2):110–116. <https://doi.org/10.1080/07408178508975280>
- Anderson RM, May RM (1992) Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford
- Anderson D, Watson R (1980) On the spread of a disease with gamma distributed latent and infectious periods. *Biometrika* 67(1):191–198. <https://doi.org/10.1093/biomet/67.1.191>
- Armbruster B, Beck E (2017) Elementary proof of convergence to the mean-field model for the SIR process. *J Math Biol* 75(2):327–339. <https://doi.org/10.1007/s00285-016-1086-1>
- Asmussen S, Nerman O, Olsson M (1996) Fitting phase-type distributions via the EM algorithm. *Scand J Stat* 23(4):419–441
- Banks HT, Catenacci J, Hu S (2013) A comparison of stochastic systems with different types of delays. *Stoch Anal Appl* 31(6):913–955. <https://doi.org/10.1080/07362994.2013.806217>
- Bindel D, Friedman M, Govaerts W, Hughes J, Kuznetsov YA (2014) Numerical computation of bifurcations in large equilibrium systems in matlab. *J Comput Appl Math* 261:232–248. <https://doi.org/10.1016/j.cam.2013.10.034>
- Blythe S, Nisbet R, Gurney W (1984) The dynamics of population models with distributed maturation periods. *Theor Popul Biol* 25(3):289–311. [https://doi.org/10.1016/0040-5809\(84\)90011-X](https://doi.org/10.1016/0040-5809(84)90011-X)

- Boese F (1989) The stability chart for the linearized Cushing equation with a discrete delay and with gamma-distributed delays. *J Math Anal Appl* 140(2):510–536. [https://doi.org/10.1016/0022-247X\(89\)90081-4](https://doi.org/10.1016/0022-247X(89)90081-4)
- Burton TA (2005) *Volterra integral and differential equations, mathematics in science and engineering*, vol 202, 2nd edn. Elsevier, Amsterdam
- Câmara De Souza D, Craig M, Cassidy T, Li J, Nekka F, Bélair J, Humphries AR (2018) Transit and lifespan in neutrophil production: implications for drug intervention. *J Pharmacokinet Pharmacodyn* 45(1):59–77. <https://doi.org/10.1007/s10928-017-9560-y>
- Campbell SA, Jessop R (2009) Approximating the stability region for a differential equation with a distributed delay. *Math Model Nat Phenom* 4(2):1–27. <https://doi.org/10.1051/mmnp/20094201>
- Champredon D, Dushoff J, Earn D (2018) Equivalence of the Erlang SEIR epidemic model and the renewal equation. *bioRxiv* <https://doi.org/10.1101/319574>
- Ciaravino G, García-Saenz A, Cabras S, Allepuz A, Casal J, García-Bocanegra I, Koeijer AD, Gubbins S, Sáez J, Cano-Terriza D, Napp S (2018) Assessing the variability in transmission of bovine tuberculosis within Spanish cattle herds. *Epidemics* 23:110–120. <https://doi.org/10.1016/j.epidem.2018.01.003>
- Clapp G, Levy D (2015) A review of mathematical models for leukemia and lymphoma. *Drug Discov Today Dis Models* 16:1–6. <https://doi.org/10.1016/j.ddmod.2014.10.002>
- Cushing JM (1994) The dynamics of hierarchical age-structured populations. *J Math Biol* 32(7):705–729. <https://doi.org/10.1007/BF00163023>
- Dhooge A, Govaerts W, Kuznetsov YA (2003) MATCONT: a MATLAB package for numerical bifurcation analysis of ODEs. *ACM Trans Math Softw* 29(2):141–164. <https://doi.org/10.1145/779359.779362>
- Diekmann O, Gyllenberg M, Metz JAJ (2017) Finite dimensional state representation of linear and nonlinear delay systems. *J Dyn Differ Equ*. <https://doi.org/10.1007/s10884-017-9611-5>
- Du P, Li MY (2014) Impact of network connectivity on the synchronization and global dynamics of coupled systems of differential equations. *Phys D Nonlinear Phenom* 286–287:32–42. <https://doi.org/10.1016/j.physd.2014.07.008>
- Ermentrout B (1987) *Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students. Software, environments and tools (Book 14)*, Society for Industrial and Applied Mathematics
- Ermentrout B (2019) XPP/XPPAUT homepage. <http://www.math.pitt.edu/~bard/xpp/xpp.html>. Accessed Apr 2019
- Fargue D (1973) Réductibilité des systèmes héréditaires à des systèmes dynamiques (régis par des équations différentielles ou aux dérivées partielles). *C R Acad Sci Paris Sér A-B* 277:B471–B473
- Feng Z, Thieme H (2000) Endemic models with arbitrarily distributed periods of infection I: fundamental properties of the model. *SIAM J Appl Math* 61(3):803–833. <https://doi.org/10.1137/S0036139998347834>
- Feng Z, Xu D, Zhao H (2007) Epidemiological models with non-exponentially distributed disease stages and applications to disease control. *Bullet Math Biol* 69(5):1511–1536. <https://doi.org/10.1007/s11538-006-9174-9>
- Feng Z, Zheng Y, Hernandez-Ceron N, Zhao H, Glasser JW, Hill AN (2016) Mathematical models of Ebola-Consequences of underlying assumptions. *Math Biosci* 277:89–107
- Fenton A, Lello J, Bonsall M (2006) Pathogen responses to host immunity: the impact of time delays and memory on the evolution of virulence. *Proc Biol Sci* 273(1597):2083–2090. <https://doi.org/10.1098/rspb.2006.3552>
- Goltser Y, Domoshnitsky A (2013) About reducing integro-differential equations with infinite limits of integration to systems of ordinary differential equations. *Adv Differ Equ* 3013 1:187. <https://doi.org/10.1186/1687-1847-2013-187>
- Guan ZH, Ling G (2018) *Dynamic analysis of genetic regulatory networks with delays*. Springer, Berlin, pp 285–309. https://doi.org/10.1007/978-3-662-55663-4_14
- Guo H, Li MY, Shuai Z (2008) A graph-theoretic approach to the method of global lyapunov functions. *Proc Am Math Soc* 136(08):2793–2802. <https://doi.org/10.1090/S0002-9939-08-09341-6>
- Gyllenberg M (2007) Mathematical aspects of physiologically structured populations: the contributions of J. A. J. Metz. *J Biol Dyn* 1(1):3–44. <https://doi.org/10.1080/17513750601032737>
- Hethcote HW, Tudor DW (1980) Integral equation models for endemic infectious diseases. *J Math Biol* 9(1):37–47. <https://doi.org/10.1007/BF00276034>

- Horváth G, Telek M (2017) BuTools 2: a rich toolbox for Markovian performance evaluation. In: Value-Tools 2016—10th EAI international conference on performance evaluation methodologies and tools, association for computing machinery, pp 137–142. <https://doi.org/10.4108/eai.25-10-2016.2266400>
- Horváth G, Reinecke P, Telek M, Wolter K (2012) Efficient Generation of PH-distributed random variates. In: Al-Begain K, Fiems D, Vincent JM (eds) Proceedings of the 19th international conference on analytical and stochastic modeling techniques and applications. Springer, Berlin, pp 271–285, https://doi.org/10.1007/978-3-642-30782-9_19
- Horváth A, Scarpa M, Telek M (2016) Phase type and matrix exponential distributions in Stochastic modeling. Springer International Publishing, Cham, pp 3–25. https://doi.org/10.1007/978-3-319-30599-8_1
- Jacquez JA, Simon CP (2002) Qualitative theory of compartmental systems with lags. *Math Biosci* 180(1):329–362. [https://doi.org/10.1016/S0025-5564\(02\)00131-1](https://doi.org/10.1016/S0025-5564(02)00131-1)
- Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc Lond Ser A Contain Pap Math Phys Character* 115(772):700–721
- Komárková Z (2012) Phase-type approximation techniques. Bachelor's Thesis, Masaryk University, <https://is.muni.cz/th/ysfsq>
- Krylova O, Earn DJD (2013) Effects of the infectious period distribution on predicted transitions in childhood disease dynamics. *J R Soc Interface*. <https://doi.org/10.1098/rsif.2013.0098>
- Krzyszanski W, Hu S, Dunlavy M (2018) Evaluation of performance of distributed delay model for chemotherapy-induced myelosuppression. *J Pharmacokinet Pharmacodyn* 45(2):329–337. <https://doi.org/10.1007/s10928-018-9575-z>
- Kurtz TG (1970) Solutions of ordinary differential equations as limits of pure jump Markov processes. *J Appl Probab* 7(1):49–58
- Kurtz TG (1971) Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J Appl Probab* 8(2):344–356. <https://doi.org/10.2307/3211904>
- Li MY, Shuai Z (2010) Global-stability problem for coupled systems of differential equations on networks. *J Differ Equ* 248(1):1–20. <https://doi.org/10.1016/j.jde.2009.09.003>
- Lin CJ, Wang L, Wolkowicz GSK (2018) An alternative formulation for a distributed delayed logistic equation. *Bull Math Biol* 80(7):1713–1735. <https://doi.org/10.1007/s11538-018-0432-4>
- Lloyd AL (2001a) Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods. *Proc R Soc Lond B Biol Sci* 268(1470):985–993. <https://doi.org/10.1098/rspb.2001.1599>
- Lloyd AL (2001b) Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor Popul Biol* 60(1):59–71. <https://doi.org/10.1006/tpbi.2001.1525>
- Lloyd AL (2009) Sensitivity of model-based epidemiological parameter estimation to model assumptions. Springer Netherlands, Dordrecht, pp 123–141. https://doi.org/10.1007/978-90-481-2313-1_6
- Ma J, Earn DJD (2006) Generality of the final size formula for an epidemic of a newly invading infectious disease. *Bull Math Biol* 68(3):679–702. <https://doi.org/10.1007/s11538-005-9047-7>
- MacDonald N (1978a) Time lags in biological models, Lecture notes in biomathematics, vol 27. Springer, Berlin. <https://doi.org/10.1007/978-3-642-93107-9>
- MacDonald N (1978b) Time lags in biological models, Lecture notes in biomathematics, vol 27, Springer, Berlin, chap Stability Analysis, pp 13–38. https://doi.org/10.1007/978-3-642-93107-9_2
- MacDonald N (1989) Biological delay systems: linear stability theory, Cambridge studies in mathematical biology, vol 8. Cambridge University Press, Cambridge
- Makroglou A, Li J, Kuang Y (2006) Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. *Appl Numer Math* 56(3):559–573. <https://doi.org/10.1016/j.apnum.2005.04.023>
- Metz JAJ, Diekmann O (eds) (1986) The dynamics of physiologically structured populations, Lecture notes in biomathematics, vol 68. Springer, Berlin. <https://doi.org/10.1007/978-3-662-13159-6>
- Metz J, Diekmann O (1991) Exact finite dimensional representations of models for physiologically structured populations. I: The abstract formulation of linear chain trickery. In: Goldstein JA, Kappel F, Schappacher W (eds) Proceedings of differential equations with applications in biology, physics, and engineering 1989, vol 133, pp 269–289
- Nisbet RM, Gurney WSC, Metz JAJ (1989) Stage structure models applied in evolutionary ecology. Springer, Berlin, pp 428–449. https://doi.org/10.1007/978-3-642-61317-3_18
- Okamura H, Dohi T (2015) Mapfit: an R-based tool for PH/MAP parameter estimation. In: Campos J, Haverkort BR (eds) Proceedings of the 12th international conference on quantitative evaluation of

- systems. Springer, New York, QEST 2015, vol 9259, pp 105–112. https://doi.org/10.1007/978-3-319-22264-6_7
- Osogami T, Harchol-Balter M (2006) Closed form solutions for mapping general distributions to quasi-minimal PH distributions. *Perform Eval* 63(6):524–552. <https://doi.org/10.1016/j.peva.2005.06.002>
- Özbay H, Bonnet C, Clairambault J (2008) Stability analysis of systems with distributed delays and application to hematopoietic cell maturation dynamics. In: 2008 47th IEEE conference on decision and control, pp 2050–2055. <https://doi.org/10.1109/CDC.2008.4738654>
- Pérez JF, Riaño G (2006) jPhase: an object-oriented tool for modeling phase-type distributions. In: Proceeding from the 2006 Workshop on tools for solving structured Markov chains, ACM, New York, NY, USA, SMCtools '06. <https://doi.org/10.1145/1190366.1190370>
- Piotrowska M, Bodnar M (2018) Influence of distributed delays on the dynamics of a generalized immune system cancerous cells interactions model. *Commun Nonlinear Sci Numer Simul* 54:389–415. <https://doi.org/10.1016/j.cnsns.2017.06.003>
- Ponosov A, Shindiapin A, Miguel JJ (2002) The W-transform links delay and ordinary differential equations. *Funct Differ Equ* 9(3–4):437–469
- Rachev ST (1991) Probability metrics and the stability of stochastic models. Wiley Series in Probability and Mathematical Statistics, Wiley, New York
- Reinecke P, Bodrog L, Danilkina A (2012a) Phase-type distributions. Springer, Berlin, pp 85–113. https://doi.org/10.1007/978-3-642-29032-9_5
- Reinecke P, Krauß T, Wolter K (2012b) Cluster-based fitting of phase-type distributions to empirical data. *Comput Math Appl* 64(12):3840–3851. <https://doi.org/10.1016/j.camwa.2012.03.016>
- Robertson SL, Henson SM, Robertson T, Cushing JM (2018) A matter of maturity: to delay or not to delay? Continuous-time compartmental models of structured populations in the literature 2000–2016. *Nat Resour Model* 31(1):e12160. <https://doi.org/10.1111/nrm.12160>
- Roussel MR (1996) The use of delay differential equations in chemical kinetics. *J Phys Chem* 100(20):8323–8330. <https://doi.org/10.1021/jp9600672>
- Smith H (2010) An introduction to delay differential equations with applications to the life sciences, vol 57. Springer, Berlin
- Smolen P, Baxter DA, Byrne JH (2000) Modeling transcriptional control in gene networks—methods, recent results, and future directions. *Bull Math Biol* 62(2):247–292. <https://doi.org/10.1006/bulm.1999.0155>
- Strogatz SH (2014) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering, 2nd edn. Studies in Nonlinearity, Westview Press
- Takashima Y, Ohtsuka T, González A, Miyachi H, Kageyama R (2011) Intronic delay is essential for oscillatory expression in the segmentation clock. *Proc Natl Acad Sci* 108(8):3300–3305. <https://doi.org/10.1073/pnas.1014418108>
- Thummler A, Buchholz P, Telek M (2006) A novel approach for phase-type fitting with the EM algorithm. *IEEE Trans Dependable Secure Comput* 3(3):245–258. <https://doi.org/10.1109/TDSC.2006.27>
- van den Driessche P, Watmough J (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci* 180(1–2):29–48. [https://doi.org/10.1016/S0025-5564\(02\)00108-6](https://doi.org/10.1016/S0025-5564(02)00108-6)
- Vogel T (1961) Systèmes déferlants, systèmes héréditaires, systèmes dynamiques. In: Proceedings of the international symposium nonlinear vibrations, IUTAM, Kiev, pp 123–130
- Vogel T (1965) Théorie des Systèmes Évolutifs. No. 22 in *Traité de physique théorique et de physique mathématique*, Gauthier-Villars, Paris
- Wang N, Han M (2016) Slow-fast dynamics of Hopfield spruce-budworm model with memory effects. *Adv Differ Equ* 1:73. <https://doi.org/10.1186/s13662-016-0804-8>
- Wearing HJ, Rohani P, Keeling MJ (2005) Appropriate models for the management of infectious diseases. *PLOS Med*. <https://doi.org/10.1371/journal.pmed.0020174>
- Wolkowicz G, Xia H, Ruan S (1997) Competition in the chemostat: a distributed delay model and its global asymptotic behavior. *SIAM J Appl Math* 57(5):1281–1310. <https://doi.org/10.1137/S0036139995289842>
- Yates CA, Ford MJ, Mort RL (2017) A multi-stage representation of cell proliferation as a Markov process. *Bull Math Biol* 79(12):2905–2928. <https://doi.org/10.1007/s11538-017-0356-4>