Abstract: Stochastic models of interacting populations have crucial roles in scientific fields such as epidemiology and ecology, yet the standard approach to extending an ordinary differential equation model to a Markov chain does not have sufficient flexibility in the mean-variance relationship to match data (e.g. [3]). A previous theory on time-homogeneous dynamics over a single arrow by [5] showed how gamma white noise could be used to construct certain over-dispersed Markov chains, leading to widely used models (e.g. [4, 9]). In this paper, we define systemic infinitesimal over-dispersion, developing theory and methodology for general time-inhomogeneous stochastic graphical models. Our approach, based on Dirichlet noise, leads to a new class of Markov models over general direct graphs. It is compatible with modern likelihood-based inference methodologies (e.g. [11, 12, 14]) and therefore we can assess how well the new models fit data. We demonstrate our methodology on a widely analyzed measles dataset, adding Dirichlet noise to a classical SEIR (Susceptible-Exposed-Infected-Recovered) model. We find that the proposed methodology has higher log-likelihood than the gamma white noise approach, and the resulting parameter estimations provide new insights into the over-dispersion of this biological system.

Keywords and phrases: General directed graph, Time-inhomogeneous stochastic processes, Infinitesimal over-dispersion, Simultaneous jumps, Extra stochasticity in data fitting, Epidemiology.

1. Introduction

In this section, we first give the background and motivations in Section 1.1 and then state our contributions in Section 1.2, followed with the organization of the paper in Section 1.3.

1.1. Background and motivations

Markov counting processes have many applications in demography, queueing theory, performance engineering, epidemiology, biology, and other areas. For example, in epidemiology, the conceptual, theoretical, and computational convenience of Markov counting processes has led to their widespread use for modeling disease transmission processes. They are building blocks for stochastic compartment models, such as the SEIR (Susceptible-Exposed-Infected-Recovered) model. In the past, a modeling hypothesis where events occur non-simultaneously has been favored. If there is no time at which two or more events occur simultaneously with probability one, a counting process is called simple otherwise.
it is called compound. Although widely acknowledged that stochastic models greatly improved model fitting compared to their deterministic counterpart, the stochastic dynamical models arising in the study of infectious disease dynamics have proved relatively recalcitrant, for the reason that extra stochasticity is to be expected in data (e.g., [3]).

There are two distinct motivations for modeling simultaneous events: First, the process in question may indeed have such occurrences; second, the process may have clusters of event times that are short compared to the scale of primary interest. Modeling disease transmission falls into this latter category, since it is easy to imagine clusters of event times, such as multiple infections caused by a sneeze on a crowded bus, while hard to imagine having exact event time data and at sufficiently fine time scales the events of becoming infected, infectious and recovered. [5] proposed a ratio-typed infinitesimal dispersion index for Markov counting processes and showed that a process has infinitesimal equi-dispersion (resp. over-dispersed) if and only if it is simple (resp. compound). [5] illustrated that the time-homogeneous infinitesimally equi-dispersed processes might be under-, equi- or over-dispersed using the integrated dispersion indices ([6]).

Stochastic epidemic models are usually time-inhomogeneous and depend on the state of other compartments/vertex for data fitting, for example, the rate of new infections in the SEIR-typed Markov chain model (equation (6) on page 332 of [4] and equation (7) on page 2583 of [5]). [4] developed the first over-dispersion methodology with the help of gamma noise for real data fitting, using the stochastic epidemic models with general transition rate functions that depend on time and state of the whole system, whereas its theoretical foundation is the infinitesimal over-dispersion generation approach on constant rates proposed in [4]. Unfortunately, the extension from constant rates to general rate functions is never trivial. It is very theoretically challenging to know the detailed properties of a stochastic integral generated by a gamma process with a general function integrand, compare to simply appending gamma noise to constant rates. Even under certain circumstances, one can theoretically achieve that, the ease to use in real applications might be lost. Therefore, not only the gap of theory and application exists between [4] and [5], but also the gap of the possible extension of gamma noise approach proposed in [5] and real applications might exists.

1.2. Our contributions

While the approach of infinitesimal over-dispersion generation with gamma noise developed in [5] has been widely used in stochastic epidemic modeling through the algorithm in [4], there is another approach in [5] that has never been used in any other literature. That is, [5] showed that infinitesimal over-dispersion can be generated by replacing the transition probabilities of one-dimensional time-homogeneous cumulative death processes with beta distributed random variables. In this paper, we give the appropriate definitions of systemic infinitesimal dispersions for graphical models with a general graph structure and general dynamics over the graph. We generalize the undeveloped beta noise approach of [5]
to construct Dirichlet noise which permits systemic over-dispersion on various subgraph structures with and without boundedness constraints. We follow the theoretical contribution by performance demonstrations on a benchmark epidemiological modeling challenge. Specifically, the contributions of the paper are four-fold:

(1) **General graph structure.** In this paper, we work on a general directed graph with general configurations over the graph and additional vertex/edge coloring, while all the preceding infinitesimal over-dispersion literature focused on one-dimensional stochastic processes over a single arrow of a small specific graphical model instead. We give an appropriate definition of systemic infinitesimal over-dispersion and provide models that can generate it over general (sub)graphs. Specifically, we provide suitable methodologies over connected arrows with the same head/tail and mutually unconnected arrows sharing common characteristics, for general dynamics with and without boundedness constraints in Sections 3.1 and 3.2, respectively.

(2) **General stochastic dynamics.** In this paper, we consider probability transition rates as general positive functions of time and state of the whole graph, while all the preceding infinitesimal over-dispersion literature considered probability transition rates being constants. A constant probability transition rate is inappropriate for stochastic epidemic modeling. In lack of theoretical support on generating infinitesimal over-dispersion for models with general transition rate functions, a theory-practice gap has been existing such as in [4] and [5]. Instead, this paper covers theory, methodology, algorithm, and data analysis. Following the theory in Section 2 and the methodology in Section 3, we provide an algorithmic Euler realization (Figure 5) and illustrate how to use it in a well-known case study in epidemiology in Section 4.

(3) **Wide applicability.** Besides the general graph structure and general dynamics over graph, wide applicability can also be seen in the following: First, only the second moment existence assumption is required which is usually satisfied in real practice; second, the software realization is simple and easy merely using beta and/or Dirichlet random variables without any computational burden; third, with our definition of systemic infinitesimal over-dispersion, users can flexibly choose those subgraphs that are appropriate to add on infinitesimal over-dispersion; fourth, the convenience of simulation from the proposed model class enables likelihood-based data fitting using simulation-based algorithms (e.g. [10], [11], [12] and [14]), thus statistically efficient inference is available for a flexible class of models, allowing scientific consideration of a broad range of hypotheses; fifth, the proposed methodology can incorporate just one additional parameter, providing a parsimonious representation of over-dispersion.

(4) **Better performance and its significance in stochastic modeling.** Stochastic models can greatly improve model fit compared to their deter-
ministic counterpart, but there are open questions remaining about the interpretation of the results. As an example, we consider estimation of the basic reproduction number, $R_0$, for measles. $R_0$ is central in epidemiological theory because it has interpretations in terms of many quantities of interest, and measles is a highly contagious infectious disease caused by measles virus which spreads easily from one person to the next through coughs and sneezes of infected people. [9] fitted an over-dispersed model with gamma noise, and found a superior model fit and also a higher maximum likelihood estimate (MLE) value of $R_0$ compared to previous estimates (see their discussion on page 276 – 278). With the same dataset, model structure, and parameter inference algorithm, in Section 4.2 we show that our Dirichlet noise model outperforms that of [9] in terms of a higher maximum log-likelihood (ML), while having an estimate of $R_0$ substantially lower than [9]. Thus, our methodology provides new scientific opportunities for modeling stochastic dynamic systems.

1.3. Organization of the paper

The rest of the paper proceeds as follows: In Section 2, we give the graph structure, configurations on the graph, dynamics over the graph, definitions of infinitesimal dispersion indices, and illustrations of the definitions through a simplified SEIR-type Markov chain model. Our methodology on generating stochastic graphical models having systemic infinitesimal over-dispersion is provided in Section 3, whose organization can be visualized in Figure 1 which is the directed graph for a two-strain infectious disease model previously used to study cholera (see, e.g., [16]). Section 3.1.1 covers outgoing arrows with the same tail, such as the brown colored arrows $(S, I_1)$, $(S, D)$, and $(S, I_2)$. Section 3.2.1 covers incoming arrows with the same head, such as the green colored arrows $(I^*_1, R)$ and $(I^*_2, R)$. Sections 3.1.2 and 3.2.2 include, but are not limited to, arrows whose heads have the same color (e.g. red color) and whose tails have the same color (e.g. purple color), such as arrows $(S_2, I^*_2)$ and $(S, I_2)$. In Section 3.3, we investigate properties of the whole graph. In Section 4, we first describe a SEIR-type Markov chain model which has been commonly used for measles data analysis in Section 4.1, and then conduct measles real data analysis on a well-tested and publicly accessible dataset using our proposed infinitesimal over-dispersion methodology and algorithm in Section 4.2. The link to data and code is provided in Acknowledgment. The notation used throughout this paper is listed in Table 2.

2. Time-inhomogeneous stochastic graphical models

In this section, we first give the graph structure and configurations on the graph in Section 2.1, dynamics over graph and their transition rates in Section 2.2, and then definitions of infinitesimal dispersion indices in Section 2.3. In Section 2.4, we illustrate the definitions covered in Sections 2.1, 2.2, and 2.3 through a SEIR-type Markov chain model.
2.1. General directed graph

A directed graph is a set of vertices connected by edges, where the edges have a direction associated with them. In this paper, we consider a finite directed graph as $G = (V,A)$, where $V$ is a set of vertices and $A$ is a set of arrows. For two vertices $v, v' \in V$, an arrow $(v,v')$ is considered to be directed from $v$ to $v'$; $v'$ is called the head and $v$ is called the tail of the arrow. In this paper, we allow a directed graph to have loops, i.e., arrows that directly connect vertices with themselves, while as in the typical setting we forbidden self-loops, i.e. arrows of the form $(v,v)$ are not contained in the directed graph. For a vertex $v \in V$, the number of heads ends adjacent to $v$ is called the indegree of $v$ and is denoted $\text{deg}^-(v)$; the number of tails ends adjacent to $v$ is called the outdegree of $v$ and is denoted $\text{deg}^+(v)$. A vertex with zero indegree is called a source and the set of all source vertices is denoted by $S_o$. A vertex with zero outdegree is called a sink and the set of all sink vertices is denoted by $S_i$. Thus,

$$S_o := \{v \in V; \text{deg}^-(v) = 0\} \quad \text{and} \quad S_i := \{v \in V; \text{deg}^+(v) = 0\}.$$
Denote given a fixed color in a finite set $C$. The graph under consideration may have arbitrary labels applied to vertices or edges. We consider vertex coloring, which colors the vertices of a graph according to a specific rule. Given that we do not restrict on the rules to apply, the vertex coloring scheme is equivalent to edge coloring scheme. We suppose that each vertex is given a fixed color in a finite set $C$.

### 2.2. Dynamics on a graph

Suppose that the dynamics of $\mathbf{X}(t) := \{X_v(t)\}_{v \in V}$ are driven by $\mathbf{N}^X \equiv \{N^X_{v,v'}(t)\}_{(v,v') \in A}$ as follows: For $v \in V$ and $t \in [0, \infty)

\[ X_v(t) = X_v(0) + \sum_{v \in N_G(v)} N^X_{v,v}(t) - \sum_{v' \in N_G(v)} N^X_{v,v'}(t). \]

We define the transition rate, under suitable conditions to suffice its existence, as follows:

\[ q(t, x, l) := \lim_{h \to 0} h^{-1} \mathbb{P}(\mathbf{N}^X(t+h) = n+1, \mathbf{X}(t+h) = x + u | \mathbf{N}^X(t) = n, \mathbf{X}(t) = x), \]

where $l = \{l_{v,v'}\}_{(v,v') \in A}$ and $u = \{u_v\}_{v \in V}$ satisfy

\[ u_v = \sum_{v \in N_G(v)} l_{v,v} - \sum_{v' \in N_G(v)} l_{v,v'}, \]

and set other transition rates to zero, based on which we define the transition rate

\[ Q(t, x, x') := \lim_{h \to 0} h^{-1} \mathbb{P}(\mathbf{X}(t+h) = x' | \mathbf{X}(t) = x). \]

Denote

\[ q(v'_iv''_i),(t,x,(k_i)i) := \sum_{l_1 \in \{l_{i,v'_i,v''_i}=k_i\}} q(t, x, l), \]

the transition rate that for each $i$, $k_i$ units transfer simultaneously through arrow $(v'_i, v''_i) \in A$, respectively. For $i$ in the set $\{1\}$,

\[ q(v'_iv''_i),(t,x,(k_i)i) = q(v'_iv''_i)(t,x,k_i) = \sum_{l_1 \in \{l_{v'_i,v''_i}=k_i\}} q(t, x, l), \]

the transition rate that $k_1$ units transfer simultaneously from vertex $v'_1$ to vertex $v''_1$ through arrow $(v'_1, v''_1) \in A$. For $i$ in the set $\{1, 2\}$,

\[ q(v'_iv''_i),(t,x,(k_i)i) = q(v'_iv''_i,v'_2v''_2)(t,x,(k_{v'_iv''_i},k_{v'_2v''_2})) = \sum_{l_1 \in \{l_{v'_i,v''_i}=k_{v'_iv''_i},l_{v'_2v''_2}=k_{v'_2v''_2}\}} q(t, x, l), \]

the transition rate that simultaneously $k_{v'_iv''_i}$ units transfer from vertex $v'_1$ to vertex $v''_1$ through arrow $(v'_1, v''_1) \in A$ and $k_{v'_2v''_2}$ units transfer from vertex $v'_2$ to vertex $v''_2$ through arrow $(v'_2, v''_2) \in A$. 

Ning and Ionides/Systemic Infinitesimal Over-dispersion
2.3. Measures of dispersion

Measures of dispersion were defined previously in the variance to mean ratio form (e.g. [7]) and the variance and mean difference form (e.g. [6]). For theoretical analysis of dispersion, these two kinds of definitions are mainly equivalent while the difference-formed definition avoids the “0/0” situation. However, the ratio-formed definition is widely used, partially due to the fact that it facilitates the dispersion comparison among different metrics and/or units. When it comes to data analysis, the over-dispersion parameter in Poisson regression uses the ratio-formed definition. Therefore in this paper, we give the definition of infinitesimal dispersion index in the ratio form as that in [5] (equation (3) on page 2574) while in the graph setting. Throughout the paper, we denote

\[ \Delta^{X_{vv}}_{vv'}(t, h) = N^{X_{vv}}_{vv'}(t + h) - N^{X_{vv}}_{vv'}(t), \quad \forall (v, v') \in A. \]

**Definition 2.1.** For arrow \((v, v') \in A\), define the infinitesimal variance

\[ [\sigma^{dX}_{vv'}(t, x)]^2 := \lim_{h \downarrow 0} h^{-1} \text{Var}[\Delta^{X_{vv'}}_{vv'}(t, h) | X(t) = x], \]

and the infinitesimal mean

\[ \mu^{dX}_{vv'}(t, x) := \lim_{h \downarrow 0} h^{-1} E[\Delta^{X_{vv'}}_{vv'}(t, h) | X(t) = x]. \]

Define the infinitesimal dispersion index as the following ratio if it exists:

\[ D^{dX}_{vv'}(t, x) := [\sigma^{dX}_{vv'}(t, x)]^2 / \mu^{dX}_{vv'}(t, x). \]

We say that with respect to arrow \((v, v') \in A\), \(X(t)\) has infinitesimal equi-dispersion at \(X(t) = x\) if \(D^{dX}_{vv'}(t, x) = 1\), has infinitesimal over-dispersion at \(X(t) = x\) if \(D^{dX}_{vv'}(t, x) > 1\), and has infinitesimal under-dispersion at \(X(t) = x\) if \(D^{dX}_{vv'}(t, x) < 1\).

Before [5], measure of dispersion refers to the integrated dispersion index, whose ratio-formed definition in the current context is the following:

**Definition 2.2.** For arrow \((v, v') \in A\), define the integrated variance

\[ [\sigma^{X}_{vv'}(t, x_0)]^2 := \text{Var}[N^{X_{vv'}}_{vv'}(t) - N^{X_{vv'}}_{vv'}(0) | X(0) = x_0], \]

and the integrated mean

\[ \mu^{X}_{vv'}(t, x_0) := E[N^{X_{vv'}}_{vv'}(t) - N^{X_{vv'}}_{vv'}(0) | X(0) = x_0]. \]

Define the integrated dispersion index as the following ratio if it exists:

\[ D^{X}_{vv'}(t, x_0) := [\sigma^{X}_{vv'}(t, x_0)]^2 / \mu^{X}_{vv'}(t, x_0). \]

We say that with respect to arrow \((v, v') \in A\), \(X(t)\) has integrated equi-dispersion if \(D^{X}_{vv'}(t, x_0) = 1\), has integrated over-dispersion if \(D^{X}_{vv'}(t, x_0) > 1\), and has integrated under-dispersion if \(D^{X}_{vv'}(t, x_0) < 1\).
Note that Definitions 2.1 and 2.2 are with respect to a specific arrow. Now we give corresponding definitions over the whole graph.

**Definition 2.3.** We say that

- $X(t)$ has systemic infinitesimal equi-dispersion at $X(t) = x$, if $D_{vv'}^X(t, x) = 1$ for all $(v, v') \in A$;
- $X(t)$ has systemic infinitesimal over-dispersion at $X(t) = x$, if $D_{vv'}^X(t, x) \geq 1$ for all $(v, v') \in A$ and there exists $(v_0, v'_0) \in A$ such that $D_{vv'_0}^X(t, x) > 1$;
- $X(t)$ has systemic infinitesimal under-dispersion at $X(t) = x$, if $D_{vv'}^X(t, x) \leq 1$ for all $(v, v') \in A$ and there exists $(v_0, v'_0) \in A$ such that $D_{vv'_0}^X(t, x) < 1$.

Note that the above definitions depend on arrow-wise variances. To explore the infinitesimal correlations over arrows’ dynamics, in the following we give the pairwise definition of infinitesimal covariance in a way that is consistent with that of infinitesimal variance in Definition 2.1.

**Definition 2.4.** For arrows $(u, u') \in A$ and $(v, v') \in A$, define the infinitesimal covariance

$$
\sigma_{u'u',vv'}^X(t, x) := \lim_{h \to 0} h^{-1} \text{Cov} [\Delta_{u'u'}^X(t, h), \Delta_{vv'}^X(t, h) \mid X(t) = x].
$$

### 2.4. Illustration through a simplified SEIR-type Markov chain model

In this subsection, we illustrate the definitions covered in Sections 2.1, 2.2 and 2.3 through a simplified SEIR-type Markov chain model. The simplifications are two-fold: First, the per capital transition rates $r_{vu}(t)$ in (2.1) and $r_{vn}(t)$ in (2.2) do not depend on state of the system; second, our analysis is arrow specific, in the way that we only consider transitions from vertex $v$ to $v'$ and do not consider other transitions to $v$. Both simplifications are necessary to illustrate the definitions of integrated mean, integrated variance, and integrated dispersion index in Definition 2.2 whose limitations and restrictions are revealed hence. We note that a general SEIR-type Markov chain model are used in the empirical data analysis in Section 4.

The SEIR model describes transitions among 6 compartments: (S) represents susceptible individuals who have not been infected yet but may experience infection later, (E) represents individuals exposed and carrying a latent infection, (I) represents infectious individuals that have been infected and are infectious to others, (R) represents recovered individuals that are no longer infectious and are immune, (B) represents the birth of individuals, and (D) represent the death of individuals. See Figure 2 for illustration. In the setting of this paper, we have $G = (V, A)$, where

- $V = \{B, S, E, I, R, D\}$,
- $S_o = \{B\}$,
- $S_i = \{D\}$,
- $A = \{(B, S), (S, E), (E, I), (I, R), (S, D), (E, D), (I, D), (R, D)\}$.
The configurations on the graph are denoted as
\[ Z(t) = z = \{ z_v \}_{v \in V} = \{ z_B, z_S, z_E, z_I, z_R, z_D \}. \]

Let the initial values be
\[ Z(0) = z^0 = \{ z_v^0 \}_{v \in V} = \{ z_B^0, z_S^0, z_E^0, z_I^0, z_R^0, z_D^0 \} \in \mathbb{N}^6. \]

We first consider a standard interpretation of the SEIR model as a Markov chain without noise on the rates. Transitions through arrow \((B, S)\) are modeled by a Poisson process with intensity function \(r_{BS}(t)\) which is assumed as a positive function uniformly continuous of \(t\). The integrated mean and the integrated variance are given by
\[ \mu_{BS}^Z(t, z_0) = \left[ \sigma_{BS}^Z(t, z_0) \right]^2 = \int_0^t r_{BS}(s)ds, \]
and then the integrated dispersion index \(D_{\text{int}}^Z(t, z_0) = 1\); the infinitesimal mean and the infinitesimal variance are given by
\[ \mu_{BS}^d(t, z) = \left[ \sigma_{BS}^d(t, z) \right]^2 = r_{BS}(t), \]
and then the infinitesimal dispersion index \(D_{\text{inf}}^Z(t, z) = 1\). Therefore, we can say that with respect to arrow \((B, S)\), \(Z(t)\) has integrated equi-dispersion and \(Z(t)\) has infinitesimal equi-dispersion at \(Z(t) = z\).

For any \((\pi, v) \in A\setminus(B, S)\), the transition rate is given by
\[ Q(t, (z_{\pi}, z_v, z_{V\setminus\{\pi, v\}})), (z_{\pi} - 1, z_v + 1, z_{V\setminus\{\pi, v\}})) = \Upsilon_{\pi v}(t, z), \]
and zero otherwise. When some components of \(Z(t)\) have non-negativity constraints, i.e. the number of individuals in each of the corresponding compartments must be non-negative at all times, modeling with unbounded processes is inappropriate. Following [4], we write
\[ \Upsilon_{\pi v}(t, z) = \tau_{\pi v}(t) z_{\pi} \mathbb{1}_{\{z_{\pi} \geq 1\}}, \tag{2.1} \]
where \(\tau_{\pi v}(t)\) is assumed as a positive function uniformly continuous of \(t\), and model the flow over arrow \((\pi, v) \in A\) by the time-inhomogeneous cumulative
death process. By [13] and Definition 2.2, the integrated mean is given by
\[
\mu_{\tau_\nu}^Z(t, z_0) = z_0^0 \left(1 - e^{-\int_0^t r_{\tau_\nu}(s)ds}\right),
\]
the integrated variance is given by
\[
[\sigma_{\tau_\nu}^Z(t, z_0)]^2 = z_0^0 e^{-\int_0^t r_{\tau_\nu}(s)ds} \left[1 - e^{-\int_0^t r_{\tau_\nu}(s)ds}\right],
\]
and then the integrated dispersion index is given by
\[
D_{\tau_\nu}^Z(t, z_0) = e^{\int_0^t r_{\tau_\nu}(s)ds}.
\]
Given \(Z(t) = z\), by l’hôpital’s rule and Definition 2.1, the infinitesimal mean and the infinitesimal variance are given by
\[
\mu_{d\tau_\nu}^Z(t, z) = [\sigma_{d\tau_\nu}^Z(t, z)]^2 = z e^{r_{\tau_\nu}(t)},
\]
and then the infinitesimal dispersion index is given by
\[
D_{d\tau_\nu}^Z(t, z) = 1.
\]
Therefore we can say that with respect to arrow \((\tau, \nu)\), \(Z(t)\) has integrated over-dispersion but \(Z(t)\) has infinitesimal equi-dispersion at \(Z(t) = z\).

1The time-inhomogeneous cumulative death process is a counting process associated with a linear death process having individual death rate \(\delta(t)\) and initial population size \(d_0 \in \mathbb{N}\), with transition rate \(q(t, m, 1) = \delta(t)(d_0 - m)1_{(m < d_0)}\) and \(q(t, m, k) = 0\) for \(k > 1\).
3. Probabilistic construction of infinitesimal over-dispersion

In this section, we probabilistically construct stochastic graphical models having systemic infinitesimal over-dispersion. Specifically, in Section 3.1.1 (resp. Section 3.2.1), we use Dirichlet noise to generate systemic infinitesimal over-dispersion over outgoing (resp. incoming) arrows with the same tail (resp. head) having bounded (resp. unbounded) constraints. In Section 3.1.2 (resp. Section 3.2.2), we use Beta noise to generate systemic infinitesimal over-dispersion over arrows that share some common characteristics having bounded (resp. unbounded) constraints. In Section 3.3, we investigate properties of the whole graph. Without loss of generality, we suppose the initial values of the dynamics over the graph are integers for notational simplicity, which will be relaxed in the empirical data analysis in Section 4 as unknown real parameters. Throughout this section, a stochastic graphical model $Z$ having systemic infinitesimal equi-dispersion is defined as follows: Conditional on $Z(t) = z$, each flow $N^Z_{v'v}$ over arrow $(v,v') \in A$ is associated with a general rate function which depends on time $t$ and state $z$, such that

$$Q(t, (z_v, z_{v'}, z_{V\setminus\{v,v'})), (z_v - 1, z_{v'} + 1, z_{V\setminus\{v,v'}))) = \Upsilon_{vv'}(t, z),$$

whose Markov chain interpretation can be specified by the infinitesimal transition probabilities:

$$P(\Delta Z_{vv'}(t,h) = 0 \mid Z(t) = z) = 1 - \Upsilon_{vv'}(t, z)h + o(h),$$
$$P(\Delta Z_{vv'}(t,h) = 1 \mid Z(t) = z) = \Upsilon_{vv'}(t, z)h + o(h),$$
$$P(\Delta Z_{vv'}(t,h) > 1 \mid Z(t) = z) = o(h),$$
$$P(\Delta Z_{vv'}(t,h) < 0 \mid Z(t) = z) = 0.$$  (3.1)

3.1. Infinitesimal over-dispersion construction with boundedness constraints

It is appropriate to model with bounded processes when modeling has non-negativity constraints, such as modeling biological population counts. In this subsection, we focus on generating stochastic graphical models having infinitesimal over-dispersion over outgoing arrows with the same tail (Section 3.1.1) and arrows that share some common characteristics (Section 3.1.2), with boundedness constraints.

3.1.1. Connected outgoing arrows modeling

In this section, we consider the case that there are multiple connected outgoing arrows of vertex $v$ such that $|N^+_G(v)| \geq 1$, where $N^+_G(v)$ is the set of vertices $v' \in V$ such that $(v,v') \in A$ and $|N^+_G(v)|$ is its cardinality. Denote $m := |N^+_G(v)|$
and \( N_{v_i}^Z(t) = \{v'_i, \ldots, v'_m\} \). The transition rate of \( N_{v_i}^Z(t) \) for \( i \in \{1, \ldots, m\} \) is given by

\[
Q(t, (z_0, v'_0), z_{v\setminus\{v'_0\}}, (z_0 - 1, v'_1 + 1, z_{v\setminus\{v'_1\}})) = r_{v'_i}(t, z) z_v \mathbb{1}_{\{z_v \geq 1\}},
\]

where \( z = (z_0, v'_0, z_{v\setminus\{v'_0\}}) \). Considering a short period \([t, t + h]\), by (3.1) the probability that one transition from vertex \( v \) to vertex \( v'_i \) for \( i \in \{1, \ldots, m\} \) is given by

\[
\mathbb{P}(\Delta_{v'_i}^Z(t, h) = 1 \mid Z(t) = z) = r_{v'_i}(t, z) z_v h + o(h).
\]

For notational convenience, denote \( \Delta_{v'_i}^Z(t, h) \) as the retain individuals at vertex \( v \). Then the joint distribution of \( \{\Delta_{v'_i}^Z(t, h) = k_i\}_{i \in \{0, \ldots, m\}} \) is given by

\[
\mathbb{P}(\{\Delta_{v'_i}^Z(t, h) = k_i\}_{i \in \{0, \ldots, m\}} \mid Z(t) = z) = \frac{\Gamma(z_v + 1)}{\prod_{i=0}^{m} \Gamma(k_i + 1)} \prod_{i=0}^{m} \left[ \tilde{\pi}_{v'_i}(t, h, z) \right]^{k_i} + o(h), \tag{3.2}
\]

where \( \Gamma(\cdot) \) is the gamma function, \( z_v \geq 1 \) and \( k_i \in \{0, 1, \ldots, z_v\} \) for \( i \in \{0, \ldots, m\} \) such that \( \sum_{i=0}^{m} k_i = z_v \). Here, for \( i \in \{1, \ldots, m\} \)

\[
\tilde{\pi}_{v'_i}(t, h, z) = \left(1 - e^{-\int_{t}^{t+h} f_{v'_i}(s, z) ds}\right) \frac{r_{v'_i}(t, z) z_v h + o(h)}{\sum_{j=1}^{m} r_{v'_j}(t, z) z_v h + o(h)} + o(h) = \left(1 - e^{-\int_{t}^{t+h} f_{v'_i}(s, z) ds}\right) \frac{r_{v'_i}(t, z) z_v h}{\sum_{j=1}^{m} r_{v'_j}(t, z) z_v h} \left(1 + o(h)\right) + o(h) = \left(1 - e^{-\int_{t}^{t+h} f_{v'_i}(s, z) ds}\right) \frac{r_{v'_i}(t, z)}{\sum_{j=1}^{m} r_{v'_j}(t, z)} \left(1 + o(h)\right) + o(h) = \left(1 - e^{-\int_{t}^{t+h} f_{v'_i}(s, z) ds}\right) \frac{r_{v'_i}(t, z)}{\sum_{j=1}^{m} r_{v'_j}(t, z)} + o(h), \tag{3.3}
\]

where we used Taylor series in the fourth equality, and

\[
\tilde{\pi}_{v'_0}(t, h, z) = 1 - \sum_{i=1}^{m} \tilde{\pi}_{v'_i}(t, h, z).
\]

Plugging (3.3) into (3.2), we can rewrite (3.2) as

\[
\mathbb{P}(\{\Delta_{v'_i}^Z(t, h) = k_i\}_{i \in \{0, \ldots, m\}} \mid Z(t) = z) = \frac{\Gamma(z_v + 1)}{\prod_{i=0}^{m} \Gamma(k_i + 1)} \prod_{i=0}^{m} \left[ \tilde{\pi}_{v'_i}(t, h, z) \right]^{k_i} + o(h),
\]
where
\[
\pi_{vv_i}(t, h, z) = \begin{cases} 
1 - e^{-\sum_{j=1}^{m} f_{t}^{i+} r_{vv_j}(v, x) ds} & \sum_{j=1}^{m} r_{vv_j}(t, x) \sum_{j=1}^{m} \pi_{vv_j}(t, h, z) \quad i \in \{1, \ldots, m\}, \\
1 - \sum_{j=1}^{m} \pi_{vv_j}(t, h, z) & i = 0.
\end{cases}
\]

The following proposition shows that a stochastic graphical model \(X\) having systemic infinitesimal over-dispersion can be generated over connected outgoing arrows \((v, v_i)_{i \in \{1, \ldots, m\}}\).

**Proposition 3.1.** Suppose that \(r_{vv_i}(t, x)\), for each \(i \in \{1, \ldots, m\}\), is a positive function that is uniformly continuous of \(t\). Further suppose that \(\{\Delta_{vv_i}(t, h) = k_i\}_{i \in \{0, \ldots, m\}}\) are jointly distributed over a short period \([t, t + h]\) as follows:

\[
P(\{\Delta_{vv_i}(t, h) = k_i\}_{i \in \{0, \ldots, m\}} \mid X(t) = x, \{\Pi_{vv_i}(t, h, x)\}_{i \in \{0, \ldots, m\}}) = \Gamma(x_v + 1) \prod_{i=0}^{m} \Gamma(k_i + 1) \prod_{i=0}^{m} (\Pi_{vv_i}(t, h, x))^{k_i} + o(h),
\]

where \(x_v \geq 1\) and \(k_i \in \{0, 1, \ldots, x_v\}\) for \(i \in \{0, \ldots, m\}\) such that \(\sum_{i=0}^{m} k_i = x_v\). Here, the family \(\{\Pi_{vv_i}(t, h, x)\}_{i \in \{0, \ldots, m\}}\) is distributed according to the Dirichlet distribution \(\text{Dir}(\{\alpha_{vv_i}(t, h, x)\}_{i \in \{0, \ldots, m\}})\) having

\[
\alpha_{vv_i}(t, h, x) = c\pi_{vv_i}(t, h, x) \quad \text{for} \quad i \in \{0, \ldots, m\}
\]

with \(c > 0\) being an inverse noise parameter and

\[
\pi_{vv_i}(t, h, x) = \begin{cases} 
1 - e^{-\sum_{j=1}^{m} f_{t}^{i+} r_{vv_j}(v, x) ds} & \sum_{j=1}^{m} r_{vv_j}(t, x) \sum_{j=1}^{m} \pi_{vv_j}(t, h, x) \quad i \in \{1, \ldots, m\}, \\
1 - \sum_{j=1}^{m} \pi_{vv_j}(t, h, x) & i = 0.
\end{cases}
\]

The following results hold:

1. For each \(i \in \{1, \ldots, m\}\), the infinitesimal mean \(\mu_{vv_i}^{X}(t, x)\) is given by

\[
\mu_{vv_i}^{X}(t, x) = x_v r_{vv_i}(t, x)
\]

and infinitesimal variance \([\sigma_{vv_i}^{X}(t, x)]^2\) is given by

\[
[\sigma_{vv_i}^{X}(t, x)]^2 = (1 + (x_v - 1)(c + 1)^{-1}) x_v r_{vv_i}(t, x).
\]

When \(x_v > 1\), \(X(t)\) has systemic infinitesimal over-dispersion at \(X(t) = x\) for connected outgoing arrows \((v, v_i)_{i \in \{1, \ldots, m\}}\); when \(x_v = 1\), \(X(t)\) has systemic infinitesimal equi-dispersion at \(X(t) = x\) for connected outgoing arrows \((v, v_i)_{i \in \{1, \ldots, m\}}\).

Furthermore, for \(i, j \in \{1, \ldots, m\}\) and \(i \neq j\), the infinitesimal covariance \(\sigma_{vv_i, vv_j}^{X}(t, x) = 0\).
(2) Denote $S$ as the set of $k_i \geq 1$ for $i \in \{1, \ldots, m\}$, i.e.,

$$S := \left\{ k_i \geq 1 \text{ for } i \in \{0, \ldots, m\}; \sum_{i=0}^{m} k_i = x_v \right\}. \quad (3.5)$$

Then we have

$$P(\{\Delta X_{vv'}(t, h) = k_i \}_{i \in \{0, \ldots, m\}}, |S| \geq 2 | X(t) = x) = o(h)$$

and

$$P(\{\Delta X_{vv'}(t, h) = k_i \}_{i \in \{0, \ldots, m\}}, |S| = 1 | X(t) = x) = \sum_{i=1}^{m} q_{vv'}(t, x, k_i) h + o(h),$$

where $|S|$ is the cardinality of $S$ and for $i \in \{1, \ldots, m\}$

$$q_{vv'}(t, x, k_i) = c \binom{x_v}{k_i} \frac{\Gamma(k_i) \Gamma(x_v - k_i + c)}{\Gamma(x_v + c)} r_{vv'}(t, x).$$

Proof. See Appendix A. 

Euler scheme for realization of the methodology proposed in Proposition 3.1 can be seen in Figure 5. A single arrow case is implied in Proposition 3.1 when $m = 1$, and we cover it in the following corollary which will be used in the proof of Proposition 3.3 and will be compared with the time-homogeneous single arrow result in [5].

Corollary 3.2. Suppose that $r_{vv'}(t, x)$ as a positive function is uniformly continuous of $t$. Further suppose that the increment of $X_{vv'}(t)$ over arrow $(v, v')$ is distributed over a short period $[t, t + h]$ as follows:

$$P(\Delta X_{vv'}(t, h) = k | X(t) = x, \Pi_{vv'}(t, h, x))$$

$$= \binom{x_v}{k} \frac{[\Pi_{vv'}(t, h, x)]^k [1 - \Pi_{vv'}(t, h, x)]^{x_v - k}}{\Gamma(x_v + c)} + o(h), \quad (3.6)$$

for $x_v \geq 1$ and $k \in \{0, 1, \ldots, x_v\}$, where $\Pi_{vv'}(t, h, x)$ is distributed according to the beta distribution $\text{Beta}(\alpha_{vv'}(t, h, x), \beta_{vv'}(t, h, x))$ having

$$\alpha_{vv'}(t, h, x) = c \pi_{vv'}(t, h, x) \quad \text{and} \quad \beta_{vv'}(t, h, x) = c(1 - \pi_{vv'}(t, h, x)),$$

with $c > 0$ being an inverse noise parameter and

$$\pi_{vv'}(t, h, x) = 1 - e^{-\int_t^{t+h} r_{vv'}(s, x) ds}.$$ 

The following results hold:
(1) The infinitesimal mean \( \mu_{vv'}(t, x) \) is given by
\[
\mu_{vv'}(t, x) = x_v r_{vv'}(t, x)
\]
and infinitesimal variance \( [\sigma_{vv'}^2(t, x)]^2 \) is given by
\[
[\sigma_{vv'}^2(t, x)]^2 = (1 + (x_v - 1)(c + 1)^{-1})x_v r_{vv'}(t, x).
\]
When \( x_v > 1 \), with respect to arrow \((v, v') \in A\), \( X(t) \) has infinitesimal over-dispersion at \( X(t) = x \); when \( x_v = 1 \), with respect to arrow \((v, v') \in A\), \( X(t) \) has infinitesimal equi-dispersion at \( X(t) = x \).

(2) For \( k \in \{1, \ldots, x_v\} \)
\[
\mathbb{P}(\Delta_{vv'}^X(t, h) = k \mid X(t) = x) = q_{vv'}(t, x, k)h + o(h), \tag{3.7}
\]
where the transition rate \( q_{vv'}(t, x, k) \) is given by
\[
q_{vv'}(t, x, k) = c \binom{x_v}{k} \frac{\Gamma(k)\Gamma(x_v - k + c)}{\Gamma(x_v + c)} r_{vv'}(t, x).
\]

Remark 3.1. For Corollary 3.2, we note the following:
(1) By the property of beta distribution, we have
\[
\mathbb{E}(\Pi_{vv'}(t, h, x)) = \pi_{vv'}(t, h, x)
\]
and
\[
\text{Var}(\Pi_{vv'}(t, h, x)) = \pi_{vv'}(t, h, x)(1 - \pi_{vv'}(t, h, x))(c + 1)^{-1}.
\]
Here, \( c \) is called an inverse noise parameter for the reason that the variance \( \text{Var}(\Pi_{vv'}(t, h, x)) \) is a decreasing function of \( c \).

(2) When \( \mu_{vv'}^X(t, x) \) is a constant for all \( t \) and \( x \), the result of Corollary 3.2 is the same as Proposition 8 on page 2581 of [5].

3.1.2. Pairwise unconnected arrows modeling

In Section 3.1.1, we provided a methodology on generating stochastic graphical models having systemic infinitesimal over-dispersion while pairwise infinitesimal covariances among dynamics of distinct outgoing arrows are zero. In practical applications, sometimes it is necessary to generate both systemic infinitesimal over-dispersion and non-zero infinitesimal covariance, such as for dynamics over arrows \((S_2, I_2^*)\) and \((S, I_2)\) whose heads have the same red color and whose tails have the same purple color for the two strain cholera model in Figure 1 (see, e.g., [16]).

We consider arrows \((v'_i, v''_{i'})_{i \in \{1, \ldots, m_1\}} \subset A\), where the colors of \( \{v'_i\}_{i \in \{1, \ldots, m_1\}} \) are \( \xi' \in \mathcal{C} \) and the colors of \( \{v''_{i'}\}_{i' \in \{1, \ldots, m_1\}} \) are \( \xi'' \in \mathcal{C} \). The transition rates are given as follows:
\[
Q(t, (z_{v'_i}, z_{v''_{i'}}, z_{V \setminus \{v'_i, v''_{i'}\}}), (z_{v'_i} - 1, z_{v''_{i'}}, z_{V \setminus \{v'_i, v''_{i'}\}}))
\]
Suppose that arrows \((v'_i, v''_i)\in\{1,\ldots,m_1\}\) are mutually not connected, and then they may be modeled as infinitesimally conditional independent cumulative death processes. Considering a short period \([t, t+h]\), for each \(i\in\{1,\ldots,m_1\}\), the increment of \(N_{v'_i,v''_i}^Z(t)\) is conditional independent distributed as follows:

\[
P(\Delta_{v'_i,v''_i}^Z(t, h) = k_i \mid Z(t) = z) = \left(\frac{z_{v'_i}}{k_i}\right)^{[\pi_{\xi,\xi'}(t, h, z)]^k_i [1 - \pi_{\xi,\xi'}(t, h, z)]^{\sum_{i=1}^{m_1} k_i} + o(h),
\]

where \(z_{v'_i} \geq 1, k_i \in \{0,1,\ldots,z_{v'_i}\}\), and

\[
\pi_{\xi,\xi'}(t, h, z) = 1 - e^{-\int_t^{t+h} r_{\xi,\xi'}(s, z) ds}.
\]

The increments of \(\{N_{v'_i,v''_i}^Z(t)\}_{i\in\{1,\ldots,m_1\}}\) are distributed as follows:

\[
P(\Delta_{v'_i,v''_i}^Z(t, h) = k_i \mid Z(t) = z) = \prod_{i=1}^{m_1} \left(\frac{z_{v'_i}}{k_i}\right)^{[\pi_{\xi,\xi'}(t, h, z)]^{k_i} [1 - \pi_{\xi,\xi'}(t, h, z)]^{\sum_{i=1}^{m_1} k_i} + o(h),
\]

where \(z_{v'_i} \geq 1\) and \(k_i \in \{0,1,\ldots,z_{v'_i}\}\) for all \(i\in\{1,\ldots,m_1\}\).

The following proposition shows that a stochastic graphical model \(\textbf{X}\) can be generated having infinitesimal over-dispersion with respect to \((v'_i, v''_i)\) for each \(i\in\{1,\ldots,m_1\}\), and positive pairwise infinitesimal covariances among dynamics of distinct arrows in \((v'_i, v''_i)\in\{1,\ldots,m_1\}\).

**Proposition 3.3.** Suppose that \(r_{\xi,\xi'}(t, \textbf{x})\) as a positive function is uniformly continuous of \(t\). Further suppose that for each \(i\in\{1,\ldots,m_1\}\) the increment of \(N_{v'_i,v''_i}^X(t)\) is conditional independent distributed over a short period \([t, t+h]\) as:

\[
P(\Delta_{v'_i,v''_i}^X(t, h) = k_i \mid X(t) = x, \Pi_{\xi,\xi'}(t, h, x)) = \left(\frac{x_{v'_i}}{k_i}\right)^{[\pi_{\xi,\xi'}(t, h, x)]^k_i [1 - \pi_{\xi,\xi'}(t, h, x)]^{\sum_{i=1}^{m_1} k_i} + o(h), \quad (3.8)
\]

where \(x_{v'_i} \geq 1\) and \(k_i \in \{0,1,\ldots,x_{v'_i}\}\) for all \(i\in\{1,\ldots,m_1\}\), and \(\Pi_{\xi,\xi'}(t, h, x)\) is distributed according to the beta distribution \(\text{Beta}(\alpha_{\xi,\xi'}(t, h, x), \beta_{\xi,\xi'}(t, h, x))\) having

\[
\alpha_{\xi,\xi'}(t, h, x) = c\pi_{\xi,\xi'}(t, h, x) \quad \text{and} \quad \beta_{\xi,\xi'}(t, h, x) = c(1 - \pi_{\xi,\xi'}(t, h, x)),
\]

with \(c > 0\) being an inverse noise parameter and

\[
\pi_{\xi,\xi'}(t, h, x) = 1 - e^{-\int_t^{t+h} r_{\xi,\xi'}(s, x) ds}.
\]

The following results hold:
(1) For each \( i \in \{1, \ldots, m_1\} \), the infinitesimal mean \( \mu_{v'_i,v'_i}(t, x) \) is given by
\[
\mu_{v'_i,v'_i}(t, x) = x_{v'_i} r_{v'_i v'_i}(t, x)
\]
and infinitesimal variance \( [\sigma_{v'_i,v'_i}(t, x)]^2 \) is given by
\[
[\sigma_{v'_i,v'_i}(t, x)]^2 = (1 + (x_{v'_i} - 1)(c + 1)^{-1}) x_{v'_i} r_{v'_i v'_i}(t, x).
\]

When \( x_{v'_i} > 1 \), \( X(t) \) has infinitesimal over-dispersion at \( X(t) = x \) with respect to arrow \((v'_i, v'_i) \in A\); when \( x_{v'_i} = 1 \), \( X(t) \) has infinitesimal equi-dispersion at \( X(t) = x \) with respect to arrow \((v'_i, v'_i) \in A\). If \( x_{v'_i} \geq 1 \) for all \( i \in \{1, \ldots, m_1\} \) and there exists \( x_{v'_k} > 1 \) where \( k \in \{1, \ldots, m_1\} \), \( X(t) \) has systemic infinitesimal over-dispersion at \( X(t) = x \) for \((v'_i, v'_i) \in \{1, \ldots, m_1\}\).

Furthermore, for \( i, j \in \{1, \ldots, m_1\} \) and \( i \neq j \), the infinitesimal covariance \( \sigma_{v'_i,v'_i,v'_j}(t, x) \) is given by
\[
\sigma_{v'_i,v'_i,v'_j}(t, x) = x_{v'_i} x_{v'_j} (c + 1)^{-1} r_{v'_i v'_j}(t, x),
\]
which is strictly positive.

(2) When \( \sum_{i=1}^{m_1} k_i \geq 1 \), we have
\[
\mathbb{P}(\{\Delta_{v'_i,v'_i}(t, h) = k_i \}_{i \in \{1, \ldots, m_1\}} \mid X(t) = x)
= q_{v'_i,v'_i}(t, x, \{k_i\}_{i \in \{1, \ldots, m_1\}}) h + o(h), \tag{3.9}
\]
where the transition rate \( q_{v'_i,v'_i}(t, x, \{k_i\}_{i \in \{1, \ldots, m_1\}}) \) is given by
\[
q_{v'_i,v'_i}(t, x, \{k_i\}_{i \in \{1, \ldots, m_1\}}) = c \prod_{i=1}^{m_1} x_{v'_i} \frac{\Gamma(\sum_{i=1}^{m_1} k_i) \Gamma(\sum_{i=1}^{m_1} x_{v'_i} - \sum_{i=1}^{m_1} k_i + c)}{\Gamma(\sum_{i=1}^{m_1} x_{v'_i} + c)} r_{v'_i v'_i}(t, x)
\]
which is strictly positive.

**Proof.** See Appendix B. \( \square \)

**Remark 3.2.** Proposition 3.3 indicates that under conditions imposed therein, infinitesimals co-jumps have positive probabilities.

### 3.2. Infinitesimal over-dispersion construction without boundedness constraints

Unbounded processes such as the pure birth process has wide applications. In this subsection, we focus on generating stochastic graphical models having infinitesimal over-dispersion over incoming arrows with the same head (Section 3.2.1) and arrows that share some common characteristics (Section 3.2.2), without boundedness constraints.
3.2.1. Connected incoming arrows modeling

In this section, we consider the case that there are multiple connected incoming arrows of vertex $u'$ such that $|N_G^{-}(u')| \geq 1$, where $N_G^{-}(u')$ is the set of vertices $u \in V$ such that $(u, u') \in A$ and $|N_G^{-}(u')|$ is its cardinality. Denote $\overline{m} := |N_G^{-}(u')|$ and $N_G^{+}(u') := \{u_1, \ldots, u_{\overline{m}}\}$. The transition probability of $N_{u_iu'}(t)$ for $i \in \{1, \ldots, \overline{m}\}$ is given by

$$Q(t, (z_{u_1}, z_{u'}, z_{V \setminus \{u, u'\}}), (z_{u_1} - 1, z_{u'} + 1, z_{V \setminus \{u, u'\}})) = r_{u_iu'}(t, z) z_{u'} 1_{\{|z_{u'}| > 0\}},$$

where $z = (z_{u_1}, z_{u'}, z_{V \setminus \{u, u'\}})$. Considering a short period $[t, t+h]$, by (3.1) the probability that one transition from vertex $u_i$ to vertex $u'$

$$\mathbb{P}(\Delta Z_{u_iu'}(t, h) = 1 \mid Z(t) = z) = r_{u_iu'}(t, z) z_{u'} h + o(h).$$

The joint distribution of increments of $\{N_{u_iu'}^{Z}(t)\}_{i \in \{1, \ldots, \overline{m}\}}$ is given by

$$\mathbb{P}(\{\Delta Z_{u_iu'}^{Z}(t, h) = k_i\}_{i \in \{1, \ldots, \overline{m}\}} \mid Z(t) = z) = \frac{\Gamma(z_{u'} + \sum_{i=1}^{\overline{m}} k_i)}{\Gamma(z_{u'}) \prod_{i=1}^{\overline{m}} \Gamma(k_i + 1)} \left[ 1 - \sum_{i=1}^{\overline{m}} \pi_{u_iu'}(t, h, z) \right]^{z_{u'} - \sum_{i=1}^{\overline{m}} \pi_{u_iu'}(t, h, z) k_i} h + o(h),$$

where $z_{u'} > 0$ and $k_i \in \{0, 1, 2, \ldots\}$ for $i \in \{1, \ldots, \overline{m}\}$. Here, for $i \in \{1, \ldots, \overline{m}\}$

$$\pi_{u_iu'}^{\overline{m}}(t, h, z) = \left(1 - e^{-\sum_{j=1}^{\overline{m}} f_j^{h} r_{u_iu'}(s, z) ds} \right) \frac{r_{u_iu'}(t, z) z_{u'} h + o(h)}{\sum_{j=1}^{\overline{m}} r_{u_iu'}(t, z) z_{u'} h + o(h)} + o(h)$$

$$= \left(1 - e^{-\sum_{j=1}^{\overline{m}} f_j^{h} r_{u_iu'}(s, z) ds} \right) \frac{r_{u_iu'}(t, z) z_{u'} h}{\sum_{j=1}^{\overline{m}} r_{u_iu'}(t, z) z_{u'} h} + o(h)$$

$$= \left(1 - e^{-\sum_{j=1}^{\overline{m}} f_j^{h} r_{u_iu'}(s, z) ds} \right) \frac{r_{u_iu'}(t, z) z_{u'} h}{\sum_{j=1}^{\overline{m}} r_{u_iu'}(t, z) z_{u'} h} \left(1 + o(h)\right)$$

$$= \left(1 - e^{-\sum_{j=1}^{\overline{m}} f_j^{h} r_{u_iu'}(s, z) ds} \right) \frac{r_{u_iu'}(t, z)}{\sum_{j=1}^{\overline{m}} r_{u_iu'}(t, z)} + o(h),$$

where we used Taylor series in the fourth equality. Plugging (3.11) into (3.10), we can rewrite (3.10) as

$$\mathbb{P}(\{\Delta Z_{u_iu'}^{Z}(t, h) = k_i\}_{i \in \{1, \ldots, \overline{m}\}} \mid Z(t) = z) = \frac{\Gamma(z_{u'} + \sum_{i=1}^{\overline{m}} k_i)}{\Gamma(z_{u'}) \prod_{i=1}^{\overline{m}} \Gamma(k_i + 1)} \left[ \pi_{u_iu'}^{\overline{m}}(t, h, z) \right]^{z_{u'} - \sum_{i=1}^{\overline{m}} \pi_{u_iu'}^{\overline{m}}(t, h, z) k_i} + o(h),$$

Ning and Ionides/Systemic Infinitesimal Over-dispersion.
where
\[
\pi_{u,u'}(t, h, z) = \begin{cases} 
1 - e^{-\sum_{j=1}^{m} \int_{t}^{t+h} r_{u,u'}(s, x)ds} \frac{r_{u,u'}(t, x)}{\sum_{j=1}^{m} r_{u,u'}(t, x)} & i \in \{1, \ldots, m\}, \\
1 - \sum_{j=1}^{m} \pi_{u,u'}(t, h, z) & i = 0.
\end{cases}
\]

The following proposition shows that a stochastic graphical model \( \mathbf{X} \) having systemic infinitesimal over-dispersion can be generated over connected incoming arrows \( (u_i, u'_i)_{i \in \{1, \ldots, m\}} \).

**Proposition 3.4.** Suppose that \( r_{u,u'}(t, x) \), for each \( i \in \{1, \ldots, m\} \), is a positive function that is uniformly continuous of \( t \). Further suppose that the increments of \( \{ \mathbf{X}_{u,u'}(t) \}_{i \in \{1, \ldots, m\}} \) are jointly distributed over a short period \([t, t+h]\) as:

\[
\begin{align*}
\mathbb{P}(\Delta_{u,u'}(t, h) = k_i)_{i \in \{1, \ldots, m\}} & \mid \mathbf{X}(t) = \mathbf{x}, \{ \Pi_{u,u'}(t, h, x) \}_{i \in \{0,1, \ldots, m\}} \\
= \frac{\Gamma(x_u + \sum_{i=1}^{m} k_i)}{\Gamma(x_u) \prod_{i=1}^{m} \Gamma(k_i + 1)} \left[ \Pi_{u,u'}(t, h, x) \right] x_u^{k_u} \prod_{i=1}^{m} \left[ \Pi_{u,u'}(t, h, x) \right]^{k_i} + o(h),
\end{align*}
\]

where \( x_u > 0 \) and \( k_i \in \{0, 1, 2, \ldots\} \) for \( i \in \{1, \ldots, m\} \). Here, the family \( \{ \Pi_{u,u'}(t, h, x) \}_{i \in \{0,1, \ldots, m\}} \) is distributed according to the Dirichlet distribution

\[
\begin{align*}
\alpha_{u,u'}(t, h, x) = c \pi_{u,u'}(t, h, x) & \text{ for } i \in \{0, 1, \ldots, m\}
\end{align*}
\]

with \( c > 0 \) being an inverse noise parameter and

\[
\pi_{u,u'}(t, h, x) = \left\{ \\
1 - e^{-\sum_{j=1}^{m} \int_{t}^{t+h} r_{u,u'}(s, x)ds} \frac{r_{u,u'}(t, x)}{\sum_{j=1}^{m} r_{u,u'}(t, x)} & i \in \{1, \ldots, m\}, \\
1 - \sum_{j=1}^{m} \pi_{u,u'}(t, h, x) & i = 0.
\right.
\]

The following results hold:

1. When \( c > 2e^{\sum_{j=1}^{m} \int_{t}^{t+h} r_{u,u'}(s, x)ds} \), for any \( i \in \{1, \ldots, m\} \), the infinitesimal mean \( \mu^{\mathbf{X}_{u,u'}}(t, x) \) is given by

\[
\mu^{\mathbf{X}_{u,u'}}(t, x) = x_u r_{u,u'}(t, x) \frac{c}{c - 1},
\]

and the infinitesimal variance \( [\sigma^{\mathbf{X}_{u,u'}}(t, x)]^2 \) is given by

\[
[\sigma^{\mathbf{X}_{u,u'}}(t, x)]^2 = x_u^2 r_{u,u'}(t, x) \frac{c}{(c - 1)(c - 2)} + x_u r_{u,u'}(t, x) \frac{c}{c - 2}.
\]

Then \( \mathbf{X}(t) \) has infinitesimal over-dispersion at \( \mathbf{X}(t) = \mathbf{x} \) with respect to each arrow of \( (u_i, u'_i)_{i \in \{1, \ldots, m\}} \), and \( \mathbf{X}(t) \) has systemic infinitesimal over-dispersion at \( \mathbf{X}(t) = \mathbf{x} \) for connected incoming arrows \( (u_i, u'_i)_{i \in \{1, \ldots, m\}} \).

Furthermore, when \( c > 2e^{\sum_{j=1}^{m} \int_{t}^{t+h} r_{u,u'}(s, x)ds} \), for \( i, j \in \{1, \ldots, m\} \) and \( i \neq j \), the infinitesimal covariance \( \sigma^{\mathbf{X}_{u_i,u'_j}}(t, x) = 0. \)
(2) Denote
\[ S := \{ k_i \geq 1 \text{ for } i \in \{1, \ldots, m\} \}. \tag{3.13} \]
Then we have
\[ P(\{ \Delta X_{uu'}(t, h) = k_i \}_{i \in \{1, \ldots, m\}}, |S| \geq 2 \mid X(t) = x) = o(h) \]
and
\[ P(\{ \Delta X_{uu'}(t, h) = k_i \}_{i \in \{1, \ldots, m\}}, |S| = 1 \mid X(t) = x) = \sum_{i=1}^{m} q_{uu'}(t, x, k_i)h + o(h), \]
where \(|S|\) is the cardinality of \( S \) and for \( i \in \{1, \ldots, m\} \)
\[ q_{uu'}(t, x, k_i) = c \frac{\Gamma(x_{u'} + \sum_{i=1}^{m} k_i)}{\Gamma(x_{u'}) \prod_{i=1}^{m} \Gamma(k_i+1)} \frac{\Gamma(x_{u'} + c)}{\Gamma(x_{u'} + \sum_{i=1}^{m} k_i + c)} r_{uu'}(t, x). \]
Proof. See Appendix C. \( \square \)

Remark 3.3. We note that users of Proposition 3.4 can simply treat \( c \) as an unknown parameter through using the parameter inference algorithm ([12]) as illustrated in Section 4.

A single arrow case is implied in Proposition 3.4 when \( m = 1 \), and we cover it in the following corollary which will be used in the proof of Proposition 3.6 and will be compared with the time-homogeneous single arrow result in [5].

Corollary 3.5. Suppose that \( r_{uu'}(t, x) \) as a positive function is uniformly continuous of \( t \). Further suppose that the increment of \( N_{uu'}(t) \) is distributed over a short period \([t, t+h]\) as:
\[ P(\Delta X_{uu'}(t, h) = k \mid X(t) = x, \Pi_{uu'}(t, h, x)) = \binom{x_{u'} + k - 1}{k} [\Pi_{uu'}(t, h, x)]^k [1 - \Pi_{uu'}(t, h, x)]^{x_{u'}-k} + o(h), \tag{3.14} \]
for \( x_{u'} > 0 \) and \( k \in \{0, 1, 2, \ldots\} \), where \( \Pi_{uu'}(t, h, x) \) is distributed according to the beta distribution Beta(\( \alpha_{uu'}(t, h, x), \beta_{uu'}(t, h, x) \)) having
\[ \alpha_{uu'}(t, h, x) = c \pi_{uu'}(t, h, x) \quad \text{and} \quad \beta_{uu'}(t, h, x) = c(1 - \pi_{uu'}(t, h, x)), \]
with \( c > 0 \) being an inverse noise parameter and
\[ \pi_{uu'}(t, h, z) = 1 - e^{-\int_{t}^{t+h} r_{uu'}(s,z)ds}. \]
The following results hold:
(1) When \( c > 2e^{\int_{t}^{t+h} r_{uu'}(s,x)ds} \), the infinitesimal mean \( \mu_{uu'}^{dx}(t,x) \) is given by

\[
\mu_{uu'}^{dx}(t,x) = x_{u'} r_{uu'}(t,x) \frac{c}{c-1},
\]

and the infinitesimal variance \( [\sigma_{uu'}^{dx}(t,x)]^2 \) is given by

\[
[\sigma_{uu'}^{dx}(t,x)]^2 = x_{u'}^2 r_{uu'}(t,x) \frac{c}{(c-1)(c-2)} + x_{u'} r_{uu'}(t,x) \frac{c}{c-2}.
\]

Then with respect to arrow \((u,u') \in A\), \(X(t)\) has infinitesimal over-dispersion at \(X(t) = x\).

(2) For \( k \in \{1,2,\ldots\} \)

\[
\mathbb{P}(\Delta_{uu'}^{x}(t,h) = k \mid X(t) = x) = q_{uu'}(t,x,k)h + o(h),
\]

(3.15)

where the transition rate \( q_{uu'}(t,x,k) \) is given by

\[
q_{uu'}(t,x,k) = c \binom{x_{u'} + k - 1}{k} \frac{\Gamma(k) \Gamma(x_{u'} + c)}{\Gamma(x_{u'} + k + c)} r_{uu'}(t,x).
\]

**Remark 3.4.** We note that the infinitesimal over-dispersion result in Corollary 3.5 holds for any \( x_{u'} > 0 \), while the gamma method proposed [5] requires that \( x_{u'} \) being an integer bigger than 1.

### 3.2.2. Pairwise unconnected arrows modeling

We focus on the same network structure as that of Section 3.1.2 while for unbounded processes. That is, we consider arrows \((u'_i,u''_i) \in \{1,\ldots,m_2\} \subset A\), where the colors of \( \{u'_i\}_{i \in \{1,\ldots,m_2\}} \) are \( \eta' \in C \) and the colors of \( \{u''_i\}_{i \in \{1,\ldots,m_2\}} \) are \( \eta'' \in C \). The transition rates are given as follows:

\[
Q(t, (z_{u'_i}, z_{u''_i}, z_V \backslash \{u'_i,u''_i\}), (z_{u'_i} - 1, z_{u''_i} + 1, z_V \backslash \{u'_i,u''_i\})) = r_{\eta'\eta''}(t,z_{u''_i}1_{z_{u''_i}>0}).
\]

Suppose that arrows \((u'_i,u''_i)_{i \in \{1,\ldots,m_2\}}\) are mutually not connected, and then they may be modeled as infinitesimally conditional independent cumulative death processes. Considering a short period \([t,t+h]\), for each \( i \in \{1,\ldots,m_2\} \), the increment of \( N_{u'_i u''_i}^{ZZ}(t) \) is conditional independent distributed as follows:

\[
\mathbb{P}(\Delta_{u'_i u''_i}^{Z}(t,h) = k_i \mid Z(t) = z) = \binom{z_{u''_i} + k_i - 1}{k_i} \left[ \pi_{\eta' \eta''}(t,h,z) \right]^{k_i} \left[ 1 - \pi_{\eta' \eta''}(t,h,z) \right]^{z_{u''_i} - k_i} + o(h),
\]

where \( z_{u''_i} > 0, k_i \in \{0,1,\ldots\} \), and

\[
\pi_{\eta' \eta''}(t,h,z) = 1 - e^{-\int_{t}^{t+h} r_{\eta' \eta''}(s,z)ds}.
\]
The increments of \( \{ N_{u_i'}^{u_i}(t) \}_{i \in \{1, \ldots, m_2 \}} \) are distributed as follows:

\[
\mathbb{P}(\sum_{i=1}^{m_2} (z_{u_i'} + k_i - 1) | \mathbf{Z}(t) = \mathbf{z})
\]

where \( z_{u_i'} > 0 \) and \( k_i \in \{0, 1, \ldots, m_2 \} \) for all \( i \in \{1, \ldots, m_2 \} \).

The following proposition shows that a stochastic graphical model \( \mathbf{X} \) can be generated having infinitesimal over-dispersion with respect to \( (u_i', u_i'') \) for each \( i \in \{1, \ldots, m_2 \} \), and positive pairwise infinitesimal covariances among dynamics of distinct arrows in \( (u_i', u_i'')_{i \in \{1, \ldots, m_2 \}} \).

**Proposition 3.6.** Suppose that \( r_{\eta \eta'}(t, \mathbf{x}) \) as a positive function is uniformly continuous of \( t \). Further suppose that for each \( i \in \{1, \ldots, m_2 \} \) the increment of \( N_{\mathbf{x}^{u_i'}}^{u_i''}(t) \) is independent distributed over a short period \( [t, t+h] \) as:

\[
\mathbb{P}(\Delta_{u_i'}^{u_i''}(t, h) = k_i \mid \mathbf{X}(t) = \mathbf{x}, \Pi_{\eta \eta'}^{u_i''}(t, h, \mathbf{x})) = \left( x_{u_i''} + k_i - 1 \right) \left[ \Pi_{\eta \eta'}^{u_i''}(t, h, \mathbf{x}) k_i [1 - \Pi_{\eta \eta'}^{u_i''}(t, h, \mathbf{x})] \right] + o(h),
\]

where \( x_{u_i''} > 0 \) and \( k_i \in \{0, 1, \ldots, m_2 \} \) for all \( i \in \{1, \ldots, m_2 \} \), and \( \Pi_{\eta \eta'}^{u_i''}(t, h, \mathbf{x}) \) is distributed according to the beta distribution \( \text{Beta}(\alpha_{\eta \eta'}^{u_i''}(t, h, \mathbf{x}), \beta_{\eta \eta'}^{u_i''}(t, h, \mathbf{x})) \) having

\[
\alpha_{\eta \eta'}^{u_i''}(t, h, \mathbf{x}) = c \pi_{\eta \eta'}^{u_i''}(t, h, \mathbf{x}) \quad \text{and} \quad \beta_{\eta \eta'}^{u_i''}(t, h, \mathbf{x}) = c(1 - \pi_{\eta \eta'}^{u_i''}(t, h, \mathbf{x}))
\]

with \( c > 0 \) being an inverse noise parameter and

\[
\pi_{\eta \eta'}^{u_i''}(t, h, \mathbf{x}) = 1 - e^{-\int_t^{t+h} r_{\eta \eta'}^{u_i''}(s, \mathbf{x}) \, ds}.
\]

The following results hold:

1. When \( c > 2e^{\int_t^{t+h} r_{\eta \eta'}^{u_i''}(s, \mathbf{x}) \, ds} \), for each \( i \in \{1, \ldots, m_2 \} \), the infinitesimal mean \( \mu_{u_i'}^{u_i''}(t, \mathbf{x}) \) is given by

\[
\mu_{u_i'}^{u_i''}(t, \mathbf{x}) = x_{u_i''} r_{\eta \eta'}^{u_i''}(t, \mathbf{x}) \frac{c}{c - 1},
\]

and the infinitesimal variance \( [\sigma_{u_i'}^{u_i''}(t, \mathbf{x})]^2 \) is given by

\[
[\sigma_{u_i'}^{u_i''}(t, \mathbf{x})]^2 = x_{u_i''}^2 r_{\eta \eta'}^{u_i''}(t, \mathbf{x}) \left( \frac{c}{c - 1}(c - 2) \right) + x_{u_i'} r_{\eta \eta'}^{u_i''}(t, \mathbf{x}) \frac{c}{c - 2}.
\]

Then \( \mathbf{X}(t) \) has infinitesimal over-dispersion at \( \mathbf{X}(t) = \mathbf{x} \) with respect to each arrow in \( (u_i', u_i'')_{i \in \{1, \ldots, m_2 \}} \), and \( \mathbf{X}(t) \) has systemic infinitesimal over-dispersion at \( \mathbf{X}(t) = \mathbf{x} \) for \( (u_i', u_i'')_{i \in \{1, \ldots, m_2 \}} \).
Furthermore, when \( c > 2e^{\int_{t}^{t+h} r_{\eta}(s, x) \, ds} \), for \( i, j \in \{1, \ldots, m_2\} \) and \( i \neq j \), the infinitesimal covariance \( \sigma_{u_{i}u_{i}'; u_{j}u_{j}'}(t, x) \) is given by
\[
\sigma_{u_{i}u_{i}'; u_{j}u_{j}'}(t, x) = x_{u_{i}'x_{u_{i}'}c(c-1)^{-1}(c-2)^{-1}r_{\eta}(t, x)},
\]
which is strictly positive.

(2) When \( \sum_{i=1}^{m_2} k_i \geq 1 \), we have
\[
P(\{\Delta X_{u_{i}u_{i}'}(t, h) = k_i\}_{i \in \{1, \ldots, m_2\}} \mid X(t) = x) = q_{u_{i}u_{i}'}(t, x, \{k_i\}_{i \in \{1, \ldots, m_2\}})h + o(h),
\]
where the transition rate \( q_{u_{i}u_{i}'}(t, x, \{k_i\}_{i \in \{1, \ldots, m_2\}}) \) is given by
\[
q_{u_{i}u_{i}'}(t, x, \{k_i\}_{i \in \{1, \ldots, m_2\}}) = c \prod_{i=1}^{m_2} \left( x_{u_{i}'} + k_i - 1 \right)^{k_i} \frac{\Gamma(\sum_{i=1}^{m_2} k_i)\Gamma(\sum_{i=1}^{m_2} x_{u_{i}'} + c)}{\Gamma(\sum_{i=1}^{m_2} x_{u_{i}'} + \sum_{i=1}^{m_2} k_i + c)} r_{\eta}(t, x),
\]
which is strictly positive.

Proof. See Appendix D. \( \square \)

Remark 3.5. Proposition 3.6 indicates that under conditions imposed therein, infinitesimal co-jumps have positive probabilities.

### 3.3. Properties of the whole graphical system

Sections 3.1 and 3.2 illustrated how to generate systemic infinitesimal over-dispersion over connected arrows with the same head/tail and mutually unconnected arrows sharing common characteristics. A natural question is whether the resulting stochastic graphical model has systemic infinitesimal over-dispersion if just some subgraphs modeled according to one of our proposed methodologies and the others do not. Another natural question is whether the resulting stochastic graphical model still carries the classical Markov chain interpretation as (3.1), since now its increments may involve increments in one or more arrows’ associated jump processes and each increment may be of size one or more. The following proposition gives answers to these two questions.

**Proposition 3.7.** For a stochastic graphical model \( X \), if there is at least one arrow’s transition dynamic modeled according to Propositions 3.1, 3.3, 3.4, and 3.6, and if the dynamics over the rest of graph have systemic infinitesimal equi-dispersion, then the following results hold:

1. \( X \) has systemic infinitesimal over-dispersion.
(2) Let $\mathcal{E}$ be the event that there is exactly one transition time occurring in the small interval $[t, t + h]$ of the stochastic graphical model, which may involve increments in one or more arrows’ associated jump processes and each increment may be of size one or more. Then the probability of $\mathcal{E}$ conditional on the configuration of the stochastic graphical model at time $t$ can be written as

$$P(\mathcal{E} \mid X(t) = x) = \lambda(t, x)h + o(h),$$

where $\lambda$ is a function that does not depend on time $h$.

Proof. (1) By Definition 2.3, infinitesimal over-dispersion with respect to a single arrow is sufficient to transfer a stochastic graphical model from having systemic infinitesimal equi-dispersion to having systemic infinitesimal over-dispersion. We finish the proof by noting that any arrow’s transition dynamic modeled according to Propositions 3.1, 3.3, 3.4 and 3.6, has infinitesimal over-dispersion.

(2) With boundedness constraints, for any vertex $v \in V$, if there are multiple connected outgoing arrows with $N^2_G(v) = \{v'_1, \ldots, v'_m\}$ and $m \geq 1$, by Proposition 3.1 we have

$$P(\{\Delta X_{v'i}^v(t, h) = k_i\} i \in \{0, 1, \ldots, m\}, |S| \geq 2 \mid X(t) = x) = o(h)$$

and

$$P(\{\Delta X_{v'i}^v(t, h) = k_i\} i \in \{0, 1, \ldots, m\}, |S| = 1 \mid X(t) = x) = \sum_{i=1}^{m} q_{v'i}(t, x, k_i)h + o(h),$$

where $x_v \geq 1$ and $k_i \in \{0, 1, \ldots, x_v\}$ for $i \in \{0, \ldots, m\}$ such that $\sum_{i=0}^{m} k_i = x_v$,

$S = \{k_i \geq 1 \text{ for } i \in \{1, \ldots, m\}\}$,

$|S|$ is the cardinality of $S$, and for $i \in \{1, \ldots, m\}$

$$q_{v'i}(t, x, k_i) = c \frac{x_v}{k_i} \frac{\Gamma(k_i) \Gamma(x_v - k_i + c)}{\Gamma(x_v + c)} r_{v'i}(t, x).$$

With boundedness constraints, for arrows $(v'_i, v''_i)_{i \in \{1, \ldots, m_1\}} \subset A$ where the colors of $(v'_i)_{i \in \{1, \ldots, m_1\}}$ are $\xi' \in \mathcal{C}$ and the colors of $(v''_i)_{i \in \{1, \ldots, m_1\}}$ are $\xi'' \in \mathcal{C}$, by Proposition 3.3 we have that

$$P(\{\Delta X_{v'i}^{v''}(t, h) = k_i\} i \in \{1, \ldots, m_1\} \mid X(t) = x) = q_{v'i}(t, x, \{k_i\} i \in \{1, \ldots, m_1\})h + o(h),$$

where $c$ is a constant.
where $x_{u'_i} \geq 1$ and $k_i \in \{0, 1, \ldots, x_{u'_i}\}$ for all $i \in \{1, \ldots, m_1\}$ such that $\sum_{i=1}^{m_1} k_i \geq 1$. Here, the transition rate $q_{\{v'_i v''_i\}_{i \in \{1, \ldots, m_1\}}} (t, x, \{k_i\}_{i \in \{1, \ldots, m_1\}})$ is given by

$$q_{\{v'_i v''_i\}_{i \in \{1, \ldots, m_1\}}} (t, x, \{k_i\}_{i \in \{1, \ldots, m_1\}}) = e \prod_{i=1}^{m_1} \left( \frac{x_{u'_i}}{k_i} \right) \frac{\Gamma(\sum_{i=1}^{m_1} k_i) \Gamma(\sum_{i=1}^{m_1} x_{u'_i} - \sum_{i=1}^{m_1} k_i + c)}{\Gamma(\sum_{i=1}^{m_1} x_{u'_i} + c)} r_{v'u''}(t, x).$$

Without boundedness constraints, for any vertex $u' \in V$, if there are multiple connected incoming arrows with $N_G(u') = \{u_1, \ldots, u_m\}$ and $m \geq 1$, by Proposition 3.4 we have

$$\mathbb{P}(\{\Delta_{u_i u'}^X(t, h) = k_i\}_{i \in \{1, \ldots, m\}}, |\mathcal{S}| \geq 2 \mid X(t) = x) = o(h),$$

and

$$\mathbb{P}(\{\Delta_{u_i u'}^X(t, h) = k_i\}_{i \in \{1, \ldots, m\}}, |\mathcal{S}| = 1 \mid X(t) = x) = \sum_{i=1}^{m} q_{u_i u'}(t, x, k_i) + o(h),$$

where $x_{u'} > 0$ and $k_i \in \{0, 1, 2, \ldots\}$ for $i \in \{1, \ldots, m\}$,

$|\mathcal{S}|$ is the cardinality of $\mathcal{S}$, and for $i \in \{1, \ldots, m\}$

$$q_{u_i u'}(t, x, k_i) = c \frac{\Gamma(x_{u'} + \sum_{i=1}^{m} k_i)}{\Gamma(x_{u'} + \sum_{i=1}^{m} k_i + 1)} \frac{\Gamma(x_{u'} + c) \Gamma(k_i)}{\Gamma(k_i + 1)} \frac{r_{u_i u'}(t, x)}{\Gamma(\sum_{i=1}^{m} k_i + c)}.$$

Without boundedness constraints, for arrows $(u'_i, u''_i)_{i \in \{1, \ldots, m_2\}} \in A$ where the colors of $(u'_i)_{i \in \{1, \ldots, m_2\}}$ are $\eta' \in \mathcal{C}$, the colors of $(u''_i)_{i \in \{1, \ldots, m_2\}}$ are $\eta'' \in \mathcal{C}$, by Proposition 3.6 we have that when $x_{u''_i} > 0$ and $k_i \in \{0, 1, \ldots\}$ for all $i \in \{1, \ldots, m_2\}$ such that $\sum_{i=1}^{m_2} k_i \geq 1$,

$$\mathbb{P}(\{\Delta^X_{u'_i u''_i}(t, h) = k_i\}_{i \in \{1, \ldots, m_2\}} \mid X(t) = x) = q_{(u'_i u''_i)}(t, x, \{k_i\}_{i \in \{1, \ldots, m_2\}}) + o(h).$$

Here, the transition rate $q_{(u'_i u''_i)}(t, x, \{k_i\}_{i \in \{1, \ldots, m_2\}})$ is given by

$$q_{(u'_i u''_i)}(t, x, \{k_i\}_{i \in \{1, \ldots, m_2\}}) = c \prod_{i=1}^{m_2} \left( \frac{x_{u''_i} + k_i - 1}{k_i} \right) \frac{\Gamma(\sum_{i=1}^{m_2} k_i) \Gamma(\sum_{i=1}^{m_2} x_{u''_i} + c)}{\Gamma(\sum_{i=1}^{m_2} x_{u''_i} + \sum_{i=1}^{m_2} k_i + c)} r_{\eta' \eta''}(t, x).$$

Given that the graph has a finite number of vertices, we complete the proof. \qed

\textit{imsart-generic ver. 2014/10/16 file: ms_06182021.tex date: June 22, 2021}
4. Application

Worldwide, measles remains a leading cause of vaccine-preventable death and disability, however global eradication of this highly infectious disease by intensive vaccination would be difficult. A fundamental class of models for measles transmission is the SEIR model, much of whose previous analysis has employed continuous-time Markov chain models where simultaneous transitions are assumed not to occur. [4] developed the first over-dispersion methodology for disease transmission with gamma noise added to the transition rates which gives an improved fit to data and widely used since then. In this section, we first describe a SEIR-type Markov chain model which has been commonly used for measles data analysis in Section 4.1. Then in Section 4.2, we conduct measles real data analysis on the same dataset as [4] and [9], which is a well-tested and publicly accessible dataset, using our methodology and algorithm, for performance comparison.

4.1. SEIR-type Markov chain model for measles

![Directed graph for the SEIR-type Markov chain models covered in Section 4.1.](image)

In this subsection, we briefly explain a SEIR-type Markov chain model and refer interested readers to [4] and [9] for detailed model interpretations. The directed graph in Fig. 3 gives a diagrammatic representation of a SEIR model, where arrows are used to indicate the possibility of transitions between vertex with labels parameterizing the transition rates. The state of the system at time t is given by the number of individuals in each vertex and is denoted as

\[ X(t) = x = \{ x_v \}_{v \in V} = \{ x_B, x_S, x_E, x_I, x_R, x_D \}. \]

The standard interpretation of Fig. 3 as a Markov chain having transition rates conditional on x is given by

\[
Q(t, (x_B, x_S, x_E, x_{R}^{(B,S)}), (x_B - 1, x_S + 1, x_{V}^{(B,S)})) = r_{BS}(t) x_B \mathbb{1}_{\{x_B > 0\}}, \\
Q(t, (x_S, x_E, x_{V}^{(S,E)}), (x_S - 1, x_E + 1, x_{V}^{(S,E)})) = r_{SE}(t, x_I) x_S \mathbb{1}_{\{x_S > 0\}}, \\
Q(t, (x_E, x_I, x_{V}^{(E,I)}), (x_E - 1, x_I + 1, x_{V}^{(E,I)})) = r_{EI} x_E \mathbb{1}_{\{x_E > 0\}}, \\
Q(t, (x_I, x_R, x_{V}^{(I,R)}), (x_I - 1, x_R + 1, x_{V}^{(I,R)})) = r_{IR} x_I \mathbb{1}_{\{x_I > 0\}}, \\
Q(t, (x_R, x_D, x_{V}^{(R,D)}), (x_R - 1, x_D + 1, x_{V}^{(R,D)})) = r_{RD} x_R \mathbb{1}_{\{x_R > 0\}}, \\
Q(t, (x_D, x_B, x_{V}^{(D,B)}), (x_D - 1, x_B + 1, x_{V}^{(D,B)})) = r_{DB} x_D \mathbb{1}_{\{x_D > 0\}}.
\]
\( Q(t, (x_I, x_R, x_{V\setminus\{I,R\}}), (x_I - 1, x_R + 1, x_{V\setminus\{I,R\}})) = r_{IR} x_I I_{\{x_I > 0\}}, \\
Q(t, (x_S, x_{D}, x_{V\setminus\{S,D\}}), (x_S - 1, x_D + 1, x_{V\setminus\{S,D\}})) = r_{SD} x_S I_{\{x_S > 0\}}, \\
Q(t, (x_E, x_{D}, x_{V\setminus\{E,D\}}), (x_E - 1, x_D + 1, x_{V\setminus\{E,D\}})) = r_{ED} x_E I_{\{x_E > 0\}}, \\
Q(t, (x_R, x_{D}, x_{V\setminus\{R,D\}}), (x_R - 1, x_D + 1, x_{V\setminus\{R,D\}})) = r_{RD} x_R I_{\{x_R > 0\}}. \\

Here, \( r_{BS}(t) \) is the per-capita rate of recruitment of susceptibles depending on known birth rates obtained via interpolation from birth records. A cohort-entry effect is also considered in calculating \( r_{BS}(t) \), to reflect the fact that a large cohort of first-year students enters the schools each fall: a fraction \( \theta_c \) of recruits into the susceptible class enter on the school admission day and the remaining fraction \((1 - \theta_c)\) enter the susceptible class continuously. We specify the force of infection as

\[
\frac{\beta(t) x_I + \iota}{N(t)},
\]

where \( \beta(t) \) is the transmission rate, \( \iota \) describes imported infectives, \( \alpha \) is a mixing parameter with \( \alpha = 1 \) corresponding to homogeneous mixing, and \( N(t) \) is a known population size obtained via interpolation from census data. Since transmission rates are closely linked to contact rates among children, which are higher during school terms, \( \beta(t) \) reflects the pattern of school terms and holidays, as follows:

\[
\beta(t) = \begin{cases}
(1 + 2(1 - p)\theta_a) \bar{\beta} & \text{during school term,} \\
(1 - 2p\theta_a) \bar{\beta} & \text{during vacation,}
\end{cases}
\]

where \( p \) is the proportion of the year taken up by school term, \( \bar{\beta} \) is the mean transmission rate, and \( \theta_a \) measures the relative effect of school holidays on transmission. For ease of interpretation, \( \bar{\beta} \) is reparameterized in terms of \( R_0 \) which is the annual average basic reproductive ratio, such that \( R_0 = \bar{\beta}/r_{IR} \), where \( r_{IR} \) is the recovery rate. Here, \( r_{EI} \) is the rate at which exposed individuals become infectious and \( r_{SD} = r_{ED} = r_{ID} = r_{RD} \) denotes a constant per capita death rate.

### 4.2. Performance comparison for real measles data

In this subsection, we demonstrate the methodology performance by analyzing measles epidemics occurring in London, England during the pre-vaccination era, with reported cases from 1950 to 1964, which is a well-tested and publicly accessible dataset. Figure 4 shows the case reports and annual birth rates for London. In order to compare with [9], we make the same decisions about which parameters to treat as fixed and which to estimate from the data. Thus, we fix \( p = 0.7589 \) in (4.1), set the delay from birth to susceptible as 4, and set the mortality rate \( r_{SD} = 1/50 = 0.02 \) per year. [9] used the over-dispersion methodology (Box 1 on page 280 of [9]) proposed in [4] for data fitting on the transition from \( S \) to \( E \), whose theory on generating over-dispersed time-homogeneous dynamics with gamma white noise is established in [5]. Therefore, we only apply
the Euler scheme on generating dynamics having systemic infinitesimal over-
dispersion over connected outgoing arrows covered in Section 3.1.1 (Figure 5),
specifically over \((S, E)\) and \((S, D)\).

The unknown model parameters in the SEIR-type Markov chain model cov-
ered in Section 4.1, are \(R_0, r_{EI}, r_{IR}, \alpha, \iota, \theta_c, \) and \(\theta_a\). To calculate the likelihood
of the data given the Markov process model, we need to add a measurement
model describing the relationship between the latent disease dynamics and the
observed case reports. We use the same measurement model as [4] and [9], which
has two more unknown parameters: reporting rate \(\rho\) and overdispersion parame-
ter \(\psi\) (see page 281 therein for a detailed description of this report measurement
process). The unknown initializations are \(X_S(0), X_E(0), X_I(0), \) and \(X_R(0)\). The
unknown infinitesimal over-dispersion model parameters are \(\sigma_{SE}\) of the gamma
white noise based approach used in [9] and \(c\) of our approach in Figure 5, respec-
tively. We implemented the same parameter inference algorithm ([12]) as [9], via
the publicly available POMP package [15]. The resulting MLE and maximum
log-likelihood (ML) are provided in Table 1. We can see that with the same
number of dimensions of unknown parameters that is the same complexity of
inference, our method has better data fitting and generates a higher ML.

The quantity \(R_0\) is central in epidemiological theory because it has interpre-
tations in terms of many quantities of interest, which include mean age of first
infection, mean susceptible fraction, exponential-phase epidemic growth rate,
and vaccination coverage required for eradication ([9]). The basic reproductive
ratio of measles is conventionally held to lie between 14 and 18 ([1]). [9] obtained
\(R_0 = 56.8\) and the likelihoods over \(R_0\) yielded approximately 95% confidence
intervals (CIs) of \((37, 60)\). Earlier than that, [2] also led to a relatively high value
\(R_0 = 29.9\) for London. [9] gave a detailed possible explanation on page 276–278,
regarding concerns on the higher MLE value of \(R_0\), while later National Insti-
(1) Set the initial value $X(0)$ and time interval $[0, T]$.
(2) Set time increment $\Delta X$; define $t_n = n \Delta X$ and $N = T/\Delta X$.
(3) FOR $n = 0$ to $N - 1$
(4) FOR each $v \in V$ with $N_G(v) = \{v'_1, \ldots, v'_m\}$
(5) Generate $\{\Pi_{vv'_i}\}_{i \in \{0, 1, \ldots, m\}}$ according to the Dirichlet distribution

$$\text{Dir}(\{\alpha_{vv'_i}\}_{i \in \{0, 1, \ldots, m\}})$$

having

$$\alpha_{vv'_i} = c \pi_{vv'_i} \quad \text{for } i \in \{1, \ldots, m\} \quad \text{and} \quad \alpha_{vv'_0} = c - \sum_{i=1}^{m} \alpha_{vv'_i},$$

where for $i \in \{1, \ldots, m\}$

$$\pi_{vv'_i} = \left(1 - e^{-\sum_{j=1}^{m} \int_{t^{k+h}}^{t^{k}} r_{vv'_i}(s, x) ds}\right) \frac{r_{vv'_i}(t, x)}{\sum_{j=1}^{m} r_{vv'_j}(t, x)}$$

and $\pi_{vv'_0} = 1 - \sum_{i=1}^{m} \pi_{vv'_i}$.

(6) Generate process increments

$\{\Delta X_{vv'_i}\}_{i \in \{0, 1, \ldots, m\}} \sim \text{Multinomial}(X_v(t_n), \{\Pi_{vv'_i}\}_{i \in \{0, 1, \ldots, m\}})$

where $\Delta X_{vv'_0}$ stands for retain individuals.

(7) Set $X_v(t_{n+1}) = \Delta X_{vv'_0} + \sum_{v \in N_G(v)} \Delta X_v$

(8) END FOR
(9) END FOR

Fig 5. Euler scheme on generating dynamics having infinitesimal over-dispersion for connected outgoing arrows covered in Section 3.1.1.
Ning and Ionides/Systemic Infinitesimal Over-dispersion

| Name                  | [9]       | Our method | Name                  | [9]       | Our method |
|-----------------------|-----------|------------|-----------------------|-----------|------------|
| Log-likelihood        | -3804.9   | -3803.2    | \( \theta_c \)       | 0.56      | 1          |
| \( R_0 \)             | 56.8      | 34.09      | \( \theta_a \)       | 0.55      | 0.48       |
| \( r_{EI} \)          | 28.9      | 52.71      | \( X_{S}(0) \)       | 0.0297    | 0.032      |
| \( r_{IR} \)          | 30.4      | 22.88      | \( X_{E}(0) \)       | 5.17e-05  | 6.99e-05   |
| \( \alpha \)          | 0.976     | 1.017      | \( X_{I}(0) \)       | 5.14e-05  | 4.52e-05   |
| \( \iota \)           | 2.9       | 55.08      | \( X_{R}(0) \)       | 0.97      | 0.968      |
| \( \rho \)            | 0.488     | 0.492      | \( c \)              | N/A       | 652.8      |
| \( \psi \)            | 0.116     | 0.118      | \( \sigma_{SE} \)    | 0.088     | N/A        |

**Table 1**
Comparison of MLE and ML generated by the same inference algorithm in [12].

Institutes of Health (NIH) quoted [8] saying: review in 2017 identified feasible measles \( R_0 \) values of 3.7–203.3. We obtained the MLE value of \( R_0 = 34.09 \) and the log-likelihoods over \( R_0 \) yielded approximately 95% CIs of (31.21, 47.37) for London (Figure 6), which is much closer to common understanding and close to the estimation obtained in [2] for London.

**Fig 6.** Log-likelihood analysis of the basic reproductive ratio, \( R_0 \). The dashed lines construct a 95% confidence interval of (31.21, 47.37) for London. This figure is calculated with the constraint that \( \alpha = 1 \), the same setting as Figure S5 of [9] for comparison purpose.

**Appendix A: Proof of Proposition 3.1**

(1) By (3.4), for \( i \in \{1, \ldots, m\} \)

\[
\mathbb{E}(\Delta_{v_i'}X(t, h) | X(t) = x) = x_v \mathbb{E}(\Pi_{v_i'}(t, h, x)) = \frac{x_v}{c} \alpha_{v_i'}(t, h, x) + o(h),
\]

\[
\text{Table 1: Comparison of MLE and ML generated by the same inference algorithm in [12].}
\]
and by law of total variance
\[ \text{Var}(\Delta_{\nu v'}(t, h) \mid X(t) = x) \]
\[ = x_v^2 \text{Var}(\Pi_{\nu v'}(t, h, x)) + x_v \mathbb{E}(\Pi_{\nu v'}(t, h, x)(1 - \Pi_{\nu v'}(t, h, x))) + o(h) \]
\[ = (x_v^2 - x_v) \frac{\alpha_{\nu v'}(t, h, x)(1 - \alpha_{\nu v'}(t, h, x))}{c^2(e + 1)} + x_v \frac{\alpha_{\nu v'}(t, h, x)}{c} \]
\[ - \frac{x_v}{c^2} \left( \alpha_{\nu v'}(t, h, x) \right)^2 + o(h). \]

Plugging
\[ \alpha_{\nu v'}(t, h, x) = c \left( 1 - e^{-\sum_{j=1}^m f_{t+h}^{x_{\nu v'}}(s, x)ds} \right) \frac{r_{\nu v'}(t, x)}{\sum_{j=1}^m r_{\nu v'}(t, x)} \]
into the above two equations, we have
\[ \mathbb{E}(\Delta_{\nu v'}^X(t, h) \mid X(t) = x) \]
\[ = x_v \left( 1 - e^{-\sum_{j=1}^m f_{t+h}^{x_{\nu v'}}(s, x)ds} \right) \frac{r_{\nu v'}(t, x)}{\sum_{j=1}^m r_{\nu v'}(t, x)} + o(h) \]
and
\[ \text{Var}(\Delta_{\nu v'}^X(t, h) \mid X(t) = x) \]
\[ = (x_v^2 - x_v)(e + 1)^{-1} \left( 1 - e^{-\sum_{j=1}^m f_{t+h}^{x_{\nu v'}}(s, x)ds} \right) \frac{r_{\nu v'}(t, x)}{\sum_{j=1}^m r_{\nu v'}(t, x)} \]
\[ \times \left( 1 - \left( 1 - e^{-\sum_{j=1}^m f_{t+h}^{x_{\nu v'}}(s, x)ds} \right) \frac{r_{\nu v'}(t, x)}{\sum_{j=1}^m r_{\nu v'}(t, x)} \right) \]
\[ + x_v \left( 1 - e^{-\sum_{j=1}^m f_{t+h}^{x_{\nu v'}}(s, x)ds} \right) \frac{r_{\nu v'}(t, x)}{\sum_{j=1}^m r_{\nu v'}(t, x)} \]
\[ - x_v \left( 1 - e^{-\sum_{j=1}^m f_{t+h}^{x_{\nu v'}}(s, x)ds} \right) \frac{r_{\nu v'}(t, x)}{\sum_{j=1}^m r_{\nu v'}(t, x)} \]
\[ \left( 1 - e^{-\sum_{j=1}^m f_{t+h}^{x_{\nu v'}}(s, x)ds} \right) \frac{r_{\nu v'}(t, x)}{\sum_{j=1}^m r_{\nu v'}(t, x)} \right)^2 + o(h). \]

Note that when \( h \) is sufficiently small, we have \( \mathbb{E}(\Delta_{\nu v'}^X(t, h) \mid X(t) = x) > 0 \) and since
\[ \left( 1 - e^{-\sum_{j=1}^m f_{t+h}^{x_{\nu v'}}(s, x)ds} \right) \frac{r_{\nu v'}(t, x)}{\sum_{j=1}^m r_{\nu v'}(t, x)} \]
\[ > \left( 1 - e^{-\sum_{j=1}^m f_{t+h}^{x_{\nu v'}}(s, x)ds} \right) \frac{r_{\nu v'}(t, x)}{\sum_{j=1}^m r_{\nu v'}(t, x)} \right)^2 \]
we also have \( \text{Var}(\Delta_{\nu v'}^X(t, h) \mid X(t) = x) > 0 \). Applying the l’hôpital’s rule,
\[ \mu_{\nu v'}^X(t, x) = x_v r_{\nu v'}(t, x) \]
By Definition 2.4, for any \( i \in \{1, \ldots, m\} \), when \( x_v > 1 \), the infinitesimal dispersion index \( D_{vv'}(t, x) > 1 \), i.e., \( X(t) \) has infinitesimal over-dispersion at \( X(t) = x \) with respect to arrow \((v, v') \in A\); when \( x_v = 1 \), the infinitesimal dispersion index \( D_{vv'}(t, x) = 1 \), i.e., \( X(t) \) has infinitesimal equi-dispersion at \( X(t) = x \) for connected outgoing arrows \((v, v') \in \{1, \ldots, m\}\). When \( x_v = 1 \), \( X(t) \) has systemic infinitesimal over-dispersion at \( X(t) = x \) for connected outgoing arrows \((v, v') \in \{1, \ldots, m\}\). By the law of total covariance, for \( i, j \in \{1, \ldots, m\} \) and \( i \neq j \) we have

\[
\text{Cov}[\Delta X_{vv'}(t, h), \Delta X_{vv'}(t, h) \mid X(t) = x] = \text{Cov}[E[\Delta X_{vv'}(t, h) \mid X(t) = x, \{\Pi_{vv'}(t, h, x)\}_{r \in \{0, \ldots, m\}}],
\]

\[
\text{Cov}[\Delta X_{vv'}(t, h), \Delta X_{vv'}(t, h) \mid X(t) = x, \{\Pi_{vv'}(t, h, x)\}_{r \in \{0, \ldots, m\}}] + E[\text{Cov}[\Delta X_{vv'}(t, h), \Delta X_{vv'}(t, h) \mid X(t) = x, \{\Pi_{vv'}(t, h, x)\}_{r \in \{0, \ldots, m\}}] = x_v^2 \text{Cov}[\Pi_{vv'}(t, h, x), \Pi_{vv'}(t, h, x)] - E[x_v \Pi_{vv'}(t, h, x) \Pi_{vv'}(t, h, x)] + o(h)
\]

\[
= (x_v^2 - x_v) \text{Cov}[\Pi_{vv'}(t, h, x), \Pi_{vv'}(t, h, x)] - x_v E[\Pi_{vv'}(t, h, x) \Pi_{vv'}(t, h, x)] + o(h)
\]

\[
= (x_v^2 - x_v) \alpha_{vv'}(t, h, x) - x_v \frac{\alpha_{vv'}(t, h, x) \alpha_{vv'}(t, h, x) c^2}{c^2} + o(h)
\]

\[
= (x_v^2 - x_v) \left(1 - e^{-\sum_{j=1}^{m} f_{i+j} r_{vv'}(s,x) ds} \right)^2 \frac{r_{vv'}(t, x) r_{vv'}(t, x)}{\left(\sum_{j=1}^{m} r_{vv'}(t, x)\right)^2} - x_v \left(1 - e^{-\sum_{j=1}^{m} f_{i+j} r_{vv'}(s,x) ds} \right)^2 \frac{r_{vv'}(t, x) r_{vv'}(t, x)}{\left(\sum_{j=1}^{m} r_{vv'}(t, x)\right)^2} + o(h).
\]

By Definition 2.4, for \( i, j \in \{1, \ldots, m\} \) and \( i \neq j \),

\[
\sigma_{\text{vv},vv'}^2(t, x) = 0.
\]

(2). By (3.4), with \( y = (y_0, y_1, \ldots, y_m) \), we have

\[
\text{P}\{\{\Delta X_{vv'}(t, h) = k_i\}_{i \in \{0, \ldots, m\}} \mid X(t) = x\}
\]

\[
= \int_0^1 \Gamma(x_v + 1) \prod_{k_i = 0}^m (y_i)^{k_i} \left\{ \frac{\Gamma(\sum_{i=0}^m \alpha_{vv'}(t, h, x))}{\prod_{i=0}^m \Gamma(\alpha_{vv'}(t, h, x))} \prod_{i=0}^m y_i^{\alpha_{vv'}(t, h, x) - 1} \right\} dy + o(h)
\]

\[
= \frac{\Gamma(x_v + 1) \prod_{k_i = 0}^m (y_i)^{k_i} \sum_{i=0}^m \alpha_{vv'}(t, h, x)}{\prod_{k_i = 0}^m \Gamma(\alpha_{vv'}(t, h, x))} \int_0^1 \prod_{i=0}^m (y_i)^{\alpha_{vv'}(t, h, x) - 1} dy + o(h)
\]
Recall that for \( k \in \mathbb{N} \),

\[
\Gamma(k + 1) = \Gamma(k) + \frac{1}{k},
\]

Hence, for \( k \), by Taylor series we have

\[
\Gamma(k + \alpha v_r'(t, h, x)) = \Gamma(k) + \frac{\alpha v_r'(t, h, x)}{k} + O(\alpha v_r'(t, h, x))^2.
\]

Plugging (A.3), (A.4), and (A.5) into (A.2), we can see that

\[
\frac{\Gamma(x - 1) \Gamma(\sum_{i=0}^{m} \alpha v_r'(t, h, x)) \prod_{i=0}^{m} \Gamma(k_i)}{\prod_{i=0}^{m} \Gamma(k_i + 1)} \prod_{i=0}^{m} \Gamma(\alpha v_r'(t, h, x)) = O(\alpha v_r'(t, h, x)).
\]

(A.2)

Recall that for \( i \in \{1, \ldots, m\} \),

\[
\alpha v_r'(t, h, x) = c \left( 1 - e^{-\sum_{j=1}^{m} \int_{t}^{t+h} r_{v_r}(s, x) ds} \right) \frac{r_{v_r}(t, x)}{\sum_{j=1}^{m} r_{v_r}(t, x)}.
\]

by Taylor series we have

\[
\alpha v_r'(t, h, x) = cr v_r'(t, x)h + O(h).
\]  

(A.3)

Hence, for \( k_i \geq 1 \) and \( i \in \{1, \ldots, m\} \),

\[
\Gamma(k_i + \alpha v_r'(t, h, x))
\]

\[
= (k_i + \alpha v_r'(t, h, x) - 1) \ldots (2 + \alpha v_r'(t, h, x) - 1) \cdot \alpha v_r'(t, h, x) \cdot \Gamma(\alpha v_r'(t, h, x))
\]

\[
= (k_i - 1) \ldots (2 - 1) \cdot \alpha v_r'(t, h, x) \cdot \Gamma(\alpha v_r'(t, h, x)) + O(\alpha v_r'(t, h, x))
\]

\[
= k_i \alpha v_r'(t, h, x) \Gamma(\alpha v_r'(t, h, x)) + O(\alpha v_r'(t, h, x)).
\]

(A.4)

Furthermore, we have that

\[
\Gamma(k_0 + \alpha v_r'(t, h, x))
\]

\[
= \left( x - \sum_{i=1}^{m} k_i - c - \sum_{i=1}^{m} \alpha v_r'(t, h, x) \right)
\]

\[
\times \left( c - \sum_{i=1}^{m} \alpha v_r'(t, h, x) \right)
\]

\[
= \left( x - \sum_{i=1}^{m} k_i - c \right) \ldots \left( c - \sum_{i=1}^{m} \alpha v_r'(t, h, x) \right) + O(\sum_{i=1}^{m} \alpha v_r'(t, h, x)) \Gamma(\alpha v_r'(t, h, x))
\]

\[
= \left( \frac{x - \sum_{i=1}^{m} k_i + c}{\Gamma(c)} \right) + O(\sum_{i=1}^{m} \alpha v_r'(t, h, x)) \Gamma(\alpha v_r'(t, h, x)).
\]

(A.5)

Plugging (A.3), (A.4), and (A.5) into (A.2), we can see that

\[
P(\{ \Delta^X_{v_r'}(t, h) = k_i \}_{i \in \{0,1,\ldots,m\}}, \ |S| \geq 2 \ | X(t) = x) = o(h).
\]

and

\[
P(\{ \Delta^X_{v_r'}(t, h) = k_i \}_{i \in \{0,1,\ldots,m\}}, \ |S| = 1 \ | X(t) = x)
\]
where $\mathcal{S}$ is the set defined in (3.5), $|\mathcal{S}|$ is the cardinality of $\mathcal{S}$, and for $i \in \{1, \ldots, m\}$

$$q_{w_i}(t, x, k_i) = c \left( \frac{x_v}{k_i} \right) \frac{\Gamma(k_i) \Gamma(x_v - k_i + c)}{\Gamma(x_v + c)} r_{w_i}(t, x).$$

### Appendix B: Proof of Proposition 3.3

(1). Since for each $i \in \{1, \ldots, m\}$, $\Delta_{v_i,v_i'}(t, h)$ is independent distributed as (3.8), by Corollary 3.2, the infinitesimal mean $\mu^{dX}_{v_i,v_i'}(t, x)$ is given by

$$\mu^{dX}_{v_i,v_i'}(t, x) = x_{v_i} r_{\xi,v_i'}(t, x)$$

and infinitesimal variance $[\sigma^{dX}_{v_i,v_i'}(t, x)]^2$ is given by

$$[\sigma^{dX}_{v_i,v_i'}(t, x)]^2 = (1 + (x_{v_i} - 1)(c + 1)^{-1}) x_{v_i} r_{\xi,v_i'}(t, x).$$

When $x_{v_i} > 1$, $X(t)$ has infinitesimal over-dispersion at $X(t) = x$ with respect to arrow $(v_i', v_i'' \in A$ and when $x_{v_i} = 1$, $X(t)$ has infinitesimal equi-dispersion at $X(t) = x$ with respect to arrow $(v_i', v_i'' \in A$. Hence, by Definition 2.3, we know that if $x_{v_i} \geq 1$ for all $i \in \{1, \ldots, m\}$ and there exists $x_{v_i} > 1$ where $k \in \{1, \ldots, m\}$, $X(t)$ has systemic infinitesimal over-dispersion at $X(t) = x$ for $(v_i', v_i'') \in \{1, \ldots, m\}$.

Given that, for $i, j \in \{1, \ldots, m\}$ and $i \neq j$, $\Delta^{X}_{v_i,v_i'}(t, h)$ and $\Delta^{X}_{v_j,v_j'}(t, h)$ are independent conditional on $X(t) = x$ and $\Pi_{\xi,v_i'}(t, h, x)$, we have

$$\text{Cov}[\Delta^{X}_{v_i,v_i'}(t, h), \Delta^{X}_{v_j,v_j'}(t, h) | X(t) = x, \Pi_{\xi,v_j'}(t, h, x)] = 0.$$
By Definition 2.4 and the l’hôpital’s rule, for \( i, j \in \{ 1, \ldots, m_1 \} \) and \( i \neq j \),

\[
\sigma_{ij}^X(t, x) = x_{ij} x_{ij} (c + 1)^{-1} r_{ij} x_{ij}(t, x),
\]

which is strictly positive.

(2) Since the increments of \( \{ N_{ij}^X(t) \}_{i \in \{ 1, \ldots, m_1 \}} \) are distributed as follows:

\[
P(\{ \Delta_{ij}^X(t, h) = k_i \}_{i \in \{ 1, \ldots, m_1 \}} \mid \mathbf{X}(t) = x, \Pi_{ij} x(t, h, x))
\]

\[
= \prod_{i=1}^{m_1} \left( \frac{x_{ij}}{k_i} \right)^{y_{ij} \sum_{i=1}^{m_1} k_i (1 - y)^{y_{ij} \sum_{i=1}^{m_1} k_i - y_{ij} \sum_{i=1}^{m_1} k_i}} \times \frac{y_{ij} \sum_{i=1}^{m_1} k_i + \alpha_{ij} x_{ij}(t, h, x)}{B(\alpha_{ij} x_{ij}(t, h, x), \beta_{ij} x_{ij}(t, h, x))} dy + o(h)
\]

\[
= \prod_{i=1}^{m_1} \left( \frac{x_{ij}}{k_i} \right) \int_0^1 \frac{y_{ij} \sum_{i=1}^{m_1} k_i + \alpha_{ij} x_{ij}(t, h, x)}{B(\alpha_{ij} x_{ij}(t, h, x), \beta_{ij} x_{ij}(t, h, x))} dy + o(h)
\]

\[
= \prod_{i=1}^{m_1} \left( \frac{x_{ij}}{k_i} \right) \frac{\Gamma(c)}{\Gamma(\alpha_{ij} x_{ij}(t, h, x)) \cdot \Gamma(\beta_{ij} x_{ij}(t, h, x))} \times \frac{\Gamma(\sum_{i=1}^{m_1} k_i + \alpha_{ij} x_{ij}(t, h, x)) \cdot \Gamma(\sum_{i=1}^{m_1} x_{ij} - \sum_{i=1}^{m_1} k_i + \beta_{ij} x_{ij}(t, h, x))}{\Gamma(\sum_{i=1}^{m_1} x_{ij} + c)} + o(h).
\]

Recall that \( \alpha_{ij} x_{ij}(t, h, x) = c \left( 1 - e^{- \int_{t}^{t+h} r_{ij} x_{ij}(s, x) ds} \right) \), by Taylor series we have

\[
\alpha_{ij} x_{ij}(t, h, x) = (c \cdot r_{ij} x_{ij}(t, x)) h + o(h).
\]

Thus, we have that when \( \sum_{i=1}^{m_1} k_i \geq 1 \),

\[
\Gamma \left( \sum_{i=1}^{m_1} k_i + \alpha_{ij} x_{ij}(t, h, x) \right)
\]

\[
= \left( \sum_{i=1}^{m_1} k_i + \alpha_{ij} x_{ij}(t, h, x) - 1 \right) \cdots \left( 2 + \alpha_{ij} x_{ij}(t, h, x) - 1 \right) \cdot \alpha_{ij} x_{ij}(t, h, x)
\]
\[ \times \Gamma(\alpha \xi'\xi''(t, h, x)) \]

\[ = \left( \sum_{i=1}^{m_1} k_i - 1 \right) \cdots (2 - 1) \cdot \alpha \xi'\xi''(t, h, x) \cdot \Gamma(\alpha \xi'\xi''(t, h, x)) + o(\alpha \xi'\xi''(t, h, x)) \]

\[ = \Gamma \left( \sum_{i=1}^{m_1} k_i \right) \alpha \xi'\xi''(t, h, x) \Gamma(\alpha \xi'\xi''(t, h, x)) + o(\alpha \xi'\xi''(t, h, x)). \]

Furthermore, we have

\[ \Gamma \left( \sum_{i=1}^{m_1} x_{i''} - \sum_{i=1}^{m_1} k_i + \beta \xi'\xi''(t, h, x) \right) \]

\[ = \left( \sum_{i=1}^{m_1} x_{i''} - \sum_{i=1}^{m_1} k_i - 1 + c - \alpha \xi'\xi''(t, h, x) \right) \cdots (c - \alpha \xi'\xi''(t, h, x)) \]

\[ \times \Gamma(\beta \xi'\xi''(t, h, x)) \]

\[ = \left[ \sum_{i=1}^{m_1} x_{i''} - \sum_{i=1}^{m_1} k_i + c \right] \cdots c + O(\alpha \xi'\xi''(t, h, x)) \] \[
\Gamma(\beta \xi'\xi''(t, h, x)) \]

\[ = \left[ \Gamma(\sum_{i=1}^{m_1} x_{i''} - \sum_{i=1}^{m_1} k_i + c) + O(\alpha \xi'\xi''(t, h, x)) \right] \Gamma(\beta \xi'\xi''(t, h, x)). \quad (B.4) \]

Plugging (B.2), (B.3) and (B.4) into (B.1), then when \( \sum_{i=1}^{m_1} k_i \geq 1 \)

\[ \mathbb{P}(\{\Delta X'_{i''}'(t, h) = k_i\}_{i \in \{1, \ldots, m_1\} } | X(t) = x) \]

\[ = q_{\{x_{i''}''\}_{i \in \{1, \ldots, m_1\} }}(t, x, \{k_i\}_{i \in \{1, \ldots, m_1\} }) + o(h), \]

where the transition rate \( q_{\{x_{i''}''\}_{i \in \{1, \ldots, m_1\} }}(t, x, \{k_i\}_{i \in \{1, \ldots, m_1\} }) \) is given by

\[ q_{\{x_{i''}''\}_{i \in \{1, \ldots, m_1\} }}(t, x, \{k_i\}_{i \in \{1, \ldots, m_1\} }) \]

\[ = c \prod_{i=1}^{m_1} \left( \frac{x_{i''}}{k_i} \right) \Gamma(\sum_{i=1}^{m_1} k_i) \frac{\Gamma(\sum_{i=1}^{m_1} x_{i''} - \sum_{i=1}^{m_1} k_i + c)}{\Gamma(\sum_{i=1}^{m_1} x_{i''} + c)} \Gamma(\xi'\xi''(t, x)), \]

which is strictly positive.

**Appendix C: Proof of Proposition 3.4**

(1) By (3.12), for \( i \in \{1, \ldots, m\} \), with \( y = (y_0, y_1, \ldots, y_m) \), we have

\[ \mathbb{E}(\Delta X'_{u'w}(t, h) | X(t) = x) \]

\[ = x_{u'} \mathbb{E} \left( (\Pi_{u'w'}(t, h, x))^{-1} \Pi_{u'w}(t, h, x) \right) + o(h) \]

\[ = x_{u'} \int_0^1 y_0^{-1} y_1 \Gamma \left( \sum_{i=0}^m \alpha_{u'w'}(t, h, x) \right) \prod_{i=0}^m \Gamma(\alpha_{u'w'}(t, h, x)) \prod_{i=0}^m y_i \alpha_{u'w'}(t, h, x) - 1 dy + o(h) \]
and law of total variance, we have

\[
\text{Var}(\Delta x_{uiw}/(t, h, x)) = \sum_{j=1}^{m} y_j \prod_{j'=1}^{i-1} y_j \prod_{j=i+1}^{m} y_j \frac{\alpha_{uiw}(t, h, x)}{\alpha_{uiw}(t, h, x) - 1} dy
\]

\[
x_u' \frac{\Gamma(\sum_{j=0}^{m} \alpha_{uiw}(t, h, x))}{\prod_{j=0}^{m} \Gamma(\alpha_{uiw}(t, h, x))} + o(h)
\]

\[
= \sum_{j=0}^{m} \Gamma(\alpha_{uiw}(t, h, x)) \prod_{j'=1}^{i-1} \Gamma(\alpha_{uiw}(t, h, x)) \prod_{j=i+1}^{m} \Gamma(\alpha_{uiw}(t, h, x))
\]

\[
x_u' \frac{\Gamma(\alpha_{uiw}(t, h, x)) \Gamma(\alpha_{uiw}(t, h, x) - 1) \Gamma(\alpha_{uiw}(t, h, x) + 1)}{\Gamma(\alpha_{uiw}(t, h, x)) \Gamma(\alpha_{uiw}(t, h, x) - 1) + o(h)}
\]

By (3.12) and law of total variance, we have

\[
\text{Var}(\Delta x_{uiw}/(t, h, x) | X(t) = x)
\]

\[
=x_u^2 \text{Var}(\Pi_{uiw}(t, h, x) - 1 \Pi_{uiw}(t, h, x))
+ x_u \text{E}[\Pi_{uiw}(t, h, x) - 2(\Pi_{uiw}(t, h, x))^2]
+ x_u \text{E}[\Pi_{uiw}(t, h, x) - 1 \Pi_{uiw}(t, h, x)] + o(h)
\]

\[
= x_u^2 \text{E}[\Pi_{uiw}(t, h, x) - 2(\Pi_{uiw}(t, h, x))^2]
+ x_u \text{E}[\Pi_{uiw}(t, h, x) - 1 \Pi_{uiw}(t, h, x)] + o(h)
\]

\[
= x_u^2 \text{E}[\Pi_{uiw}(t, h, x) - 2(\Pi_{uiw}(t, h, x))^2]
+ x_u \text{E}[\Pi_{uiw}(t, h, x) - 1 \Pi_{uiw}(t, h, x)] + o(h)
\]

\[
= x_u^2 \frac{\Gamma(\alpha_{uiw}(t, h, x) - 2) \Gamma(\alpha_{uiw}(t, h, x) + 2)}{\Gamma(\alpha_{uiw}(t, h, x)) \Gamma(\alpha_{uiw}(t, h, x))}
\]

\[
- x_u \left( \frac{\Gamma(\alpha_{uiw}(t, h, x) - 1) \Gamma(\alpha_{uiw}(t, h, x) + 1)}{\Gamma(\alpha_{uiw}(t, h, x)) \Gamma(\alpha_{uiw}(t, h, x))} \right)^2
\]

\[
+ x_u \left( \frac{\Gamma(\alpha_{uiw}(t, h, x) - 1) \Gamma(\alpha_{uiw}(t, h, x) + 1)}{\Gamma(\alpha_{uiw}(t, h, x)) \Gamma(\alpha_{uiw}(t, h, x))} \right) + o(h)
\]

\[
= x_u^2 \left( \frac{\alpha_{uiw}(t, h, x) + 1}{\alpha_{uiw}(t, h, x) - 1} \right)^2
+ x_u \left( \frac{\alpha_{uiw}(t, h, x)}{\alpha_{uiw}(t, h, x) - 1} \right) + o(h).
\]
into equations (C.1) and (C.2), we have

$$E(\Delta_{u,t}^X(t, h) \mid X(t) = x)$$

$$= x_u \frac{c}{ce - \sum_{j=1}^{\infty} f_{\mu_j} r_{u_j u'}(s,x) ds} \frac{r_{u_j u'}(t,x)}{\sum_{j=1}^{\infty} r_{u_j u'}(t,x)} + o(h) \quad (C.5)$$

and

$$\text{Var}(\Delta_{u,t}^X(t, h) \mid X(t) = x)$$

$$= (x_u^2 + x_u^2) \frac{c}{ce - \sum_{j=1}^{\infty} f_{\mu_j} r_{u_j u'}(s,x) ds} \frac{r_{u_j u'}(t,x)}{\sum_{j=1}^{\infty} r_{u_j u'}(t,x)} + 1$$

$$\times c \left( 1 - e^{-\sum_{j=1}^{\infty} f_{\mu_j} r_{u_j u'}(s,x) ds} \right) \frac{r_{u_j u'}(t,x)}{\sum_{j=1}^{\infty} r_{u_j u'}(t,x)} + 1$$

$$- x_u^2 \frac{c}{ce - \sum_{j=1}^{\infty} f_{\mu_j} r_{u_j u'}(s,x) ds} \frac{r_{u_j u'}(t,x)}{\sum_{j=1}^{\infty} r_{u_j u'}(t,x)}$$

$$+ x_u^2 \frac{c}{ce - \sum_{j=1}^{\infty} f_{\mu_j} r_{u_j u'}(s,x) ds} + o(h). \quad (C.6)$$

Note that when $h$ is sufficiently small, $c > 2e^{\sum_{j=1}^{\infty} f_{\mu_j} r_{u_j u'}(s,x) ds}$ suffices that $E(\Delta_{u,t}^X(t, h) \mid X(t) = x) > 0$, and together with the fact that

$$\left( 1 - e^{-\sum_{j=1}^{\infty} f_{\mu_j} r_{u_j u'}(s,x) ds} \right) \frac{r_{u_j u'}(t,x)}{\sum_{j=1}^{\infty} r_{u_j u'}(t,x)} + 1$$

$$\frac{c}{ce - \sum_{j=1}^{\infty} f_{\mu_j} r_{u_j u'}(s,x) ds} + 2$$

$$> c \left( 1 - e^{-\sum_{j=1}^{\infty} f_{\mu_j} r_{u_j u'}(s,x) ds} \right) \frac{r_{u_j u'}(t,x)}{\sum_{j=1}^{\infty} r_{u_j u'}(t,x)}$$

we also have $\text{Var}(\Delta_{u,t}^X(t, h) \mid X(t) = x) > 0$.

By the l'hôpital’s rule, equations (C.5) and (C.6), and Definition 2.1, we have

$$\mu_{u,t}^X(t, x) = x_u r_{u_j u'}(t, x) \frac{c}{c - 1},$$

$$[\sigma_{u,t}^X(t, x)]^2 = x_u^2 r_{u_j u'}(t, x) \frac{c}{(c - 1)(c - 2)} + x_u r_{u_j u'}(t, x) \frac{c}{c - 2}.$$
and
\[
x_\alpha^2 r_{u,u'}(t,x) \frac{c}{(c-1)(c-2)} > 0
\]
we have \(D^\infty_{u,u'}(t,x) > 1\) always, i.e., \(X(t)\) has infinitesimal over-dispersion at \(X(t) = x\) with respect to each arrow in \((u_i, u'_i)_{i \in \{1, \ldots, m\}}\). By Definition 2.3, \(X(t)\) has systemic infinitesimal over-dispersion at \(X(t) = x\) for connected incoming arrows \((u_i, u'_i)_{i \in \{1, \ldots, m\}}\).

By the law of total covariance, for \(i, j \in \{1, \ldots, m\}\) and \(i \neq j\) we have
\[
\text{Cov}[\Delta_{u,u'}(t,h), \Delta_{u,u'}(t,h) | X(t) = x]
= \text{Cov}[E[\Delta_{u,u'}(t,h) | X(t) = x, \{\Pi_{u,u'}(t,h,x)\}_{i \in \{0,1,\ldots,m\}}],
E[\Delta_{u,u'}(t,h) | X(t) = x, \{\Pi_{u,u'}(t,h,x)\}_{i \in \{0,1,\ldots,m\}}]] + E[\text{Cov}[\Delta_{u,u'}(t,h), \Delta_{u,u'}(t,h) | X(t) = x, \{\Pi_{u,u'}(t,h,x)\}_{i \in \{0,1,\ldots,m\}}]] + o(h)
\]
\[
= x_\alpha \text{Cov}[(\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x), (\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x)] + E[x_{u'}(\Pi_{u,u'}(t,h,x))^{-2}\Pi_{u,u'}(t,h,x)\Pi_{u,u'}(t,h,x) + o(h)
\]
\[
= x_\alpha \text{Cov}[(\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x), (\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x)] + E[x_{u'}(\Pi_{u,u'}(t,h,x))^{-2}\Pi_{u,u'}(t,h,x)\Pi_{u,u'}(t,h,x) + o(h)
\]
\[
= x_\alpha (x_{u'}^2)E[(\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x), (\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x)]
+ E[(\Pi_{u,u'}(t,h,x))^{-2}\Pi_{u,u'}(t,h,x)\Pi_{u,u'}(t,h,x) + o(h)
\]
\[
= (x_\alpha^2 + x_{u'}^2)E[(\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x), (\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x)]
+ x_{u'}^2(\Pi_{u,u'}(t,h,x))^{-2}\Pi_{u,u'}(t,h,x)\Pi_{u,u'}(t,h,x) + o(h)
\]
\[
= (x_\alpha^2 + x_{u'}^2)E[(\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x), (\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x)]
+ x_{u'}^2(\Pi_{u,u'}(t,h,x))^{-2}\Pi_{u,u'}(t,h,x)\Pi_{u,u'}(t,h,x) + o(h)
\]
\[
= (x_\alpha^2 + x_{u'}^2)E[(\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x), (\Pi_{u,u'}(t,h,x))^{-1}\Pi_{u,u'}(t,h,x)]
+ x_{u'}^2(\Pi_{u,u'}(t,h,x))^{-2}\Pi_{u,u'}(t,h,x)\Pi_{u,u'}(t,h,x) + o(h).
\]
(C.7)

Note that
\[
E[(\Pi_{u,u'}(t,h,x))^{-2}\Pi_{u,u'}(t,h,x)\Pi_{u,u'}(t,h,x)]
= \int_0^1 y_0^{-2}y_1 y_2 \frac{\Gamma(\sum_{i=0}^{m} \alpha_{u_i,u'}(t,h,x))}{\prod_{i=0}^{m} \Gamma(\alpha_{u_i,u'}(t,h,x))} \prod_{i=0}^{m} y_i \alpha_{u_i,u'}(t,h,x)^{-1} dy + o(h)
\]
\[
= \frac{\Gamma(\sum_{i=0}^{m} \alpha_{u_i,u'}(t,h,x))}{\prod_{i=0}^{m} \Gamma(\alpha_{u_i,u'}(t,h,x))} \prod_{i=0}^{m} y_i \alpha_{u_i,u'}(t,h,x)^{-1} dy + o(h)
\]
\[
= \Gamma(\alpha_{u,u'}(t,h,x) - 2)\Gamma(\alpha_{u,u'}(t,h,x) + 1)\Gamma(\alpha_{u,u'}(t,h,x) + 1)
\]
\[
\times \frac{\Gamma(\sum_{i=0}^{m} \alpha_{u_i,u'}(t,h,x))}{\prod_{i=0}^{m} \Gamma(\alpha_{u_i,u'}(t,h,x))} \prod_{i=0}^{m} \Gamma(\alpha_{u_i,u'}(t,h,x)) \Gamma(\sum_{i=0}^{m} \alpha_{u_i,u'}(t,h,x)) + o(h)
\]
\[
= \frac{\Gamma(\alpha_{u,u'}(t,h,x) - 2)\Gamma(\alpha_{u,u'}(t,h,x) + 1)\Gamma(\alpha_{u,u'}(t,h,x) + 1)}{\alpha_{u,u'}(t,h,x) \alpha_{u,u'}(t,h,x) \alpha_{u,u'}(t,h,x)} + o(h).
\]
(C.8)
Plugging (C.9) into (C.7), we have
\[
\mathbb{E}[(\Pi_{u|t'}(t, h, x))^{-2}\Pi_{u|t'}(t, h, x)\Pi_{u|t'}(t, h, x)]
= \frac{c^2 \left(1 - e^{-\sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds} \right)^2 r_{u|t'}(t, x) r_{u|t'}(t, x)}{(ce - \sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds - 1)(ce - \sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds - 2)} + o(h). \tag{C.9}
\]

Plugging (C.9) into (C.7), we have
\[
\text{Cov}[\Delta_{u|t'}(t, h), \Delta_{u|t'}(t, h) | X(t) = x]
= (x_u'^2 + x_u') \frac{c^2 \left(1 - e^{-\sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds} \right)^2 r_{u|t'}(t, x) r_{u|t'}(t, x)}{(ce - \sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds - 1)(ce - \sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds - 2)}
- x_u' \frac{c \left(1 - e^{-\sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds} \right) \sum_{j=1}^{m} r_{u|t'}(t, x)}{(ce - \sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds - 1)}
\times \frac{c \left(1 - e^{-\sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds} \right) \sum_{j=1}^{m} r_{u|t'}(t, x)}{(ce - \sum_{j=1}^{m} \int_{t}^{t+h} r_{u|t'}(s, x) ds - 1) + o(h). \tag{C.10}
\]

By Definition 2.4, for \(i, j \in \{1, \ldots, m\}\) and \(i \neq j\)
\[
\sigma_{u|t'}^{(i, j)}(t, x) = 0.
\]

(2) By (3.12), with \(y = (y_0, y_1, \ldots, y_m)\), we have
\[
\mathbb{P}(\{\Delta_{u|t'}(t, h) = k_i\}_{i \in \{1, \ldots, m\}} | X(t) = x)
= \int_0^1 \frac{\Gamma(x_u' + \sum_{i=1}^{m} k_i)}{\Gamma(x_u')} \prod_{i=1}^{m} \Gamma(k_i + 1) \prod_{i=1}^{m} \left[y_i\right]^{\alpha_{u|t'}(t, h, x)} \frac{\Gamma(\sum_{i=0}^{m} \alpha_{u|t'}(t, h, x))}{\prod_{i=0}^{m} \Gamma(\alpha_{u|t'}(t, h, x))} \times \prod_{i=0}^{m} \frac{y_i^{\alpha_{u|t'}(t, h, x) - 1} dy + o(h)}{y_i}
= \frac{\Gamma(x_u' + \sum_{i=1}^{m} k_i)}{\Gamma(x_u')} \prod_{i=1}^{m} \Gamma(k_i + 1) \prod_{i=0}^{m} \Gamma(\alpha_{u|t'}(t, h, x))
\times \int_0^1 \left[y_0\right]^{x_u' + \alpha_{u|t'}(t, h, x)} \prod_{i=1}^{m} \left[y_i\right]^{k_i + \alpha_{u|t'}(t, h, x) - 1} dy + o(h)
= \frac{\Gamma(x_u' + \sum_{i=1}^{m} k_i)}{\Gamma(x_u')} \prod_{i=1}^{m} \Gamma(k_i + 1) \prod_{i=0}^{m} \Gamma(\alpha_{u|t'}(t, h, x))
\times \frac{\Gamma(x_u' + \alpha_{u|t'}(t, h, x)) \prod_{i=1}^{m} \Gamma(k_i + \alpha_{u|t'}(t, h, x))}{\Gamma(x_u' + \sum_{i=1}^{m} k_i + \sum_{i=0}^{m} \alpha_{u|t'}(t, h, x)) + o(h)}.
by Taylor series we have

\[ H(k_i + \alpha_{u,w}(t, h, x)) = (k_i + \alpha_{u,w}(t, h, x) - 1) \cdots (2 + \alpha_{u,w}(t, h, x) - 1) \cdot \alpha_{u,w}(t, h, x) \times \Gamma(\alpha_{u,w}(t, h, x)) \]

Thus, for \( k_i \geq 1 \) and \( i \in \{1, \ldots, m\} \),

\[ \Gamma(k_i + \alpha_{u,w}(t, h, x)) = (k_i - 1) \cdots (2 - 1) \cdot \alpha_{u,w}(t, h, x) \cdot \Gamma(\alpha_{u,w}(t, h, x)) + o(\alpha_{u,w}(t, h, x)) \]

Furthermore, we have

\[ \Gamma(x_{u'} + \alpha_{u,w}(t, h, x)) = \left( x_{u'} - 1 + c - \sum_{i=1}^{m} \alpha_{u,w}(t, h, x) \right) \cdots \left( c - \sum_{i=1}^{m} \alpha_{u,w}(t, h, x) \right) \Gamma(\alpha_{u,w}(t, h, x)) \]

\[ = \left[ (x_{u'} - 1 + c) \cdots c + \mathcal{O} \left( \sum_{i=1}^{m} \alpha_{u,w}(t, h, x) \right) \right] \Gamma(\alpha_{u,w}(t, h, x)) \]

\[ = \left[ \frac{\Gamma(x_{u'} + c)}{\Gamma(c)} + \mathcal{O} \left( \sum_{i=1}^{m} \alpha_{u,w}(t, h, x) \right) \right] \Gamma(\alpha_{u,w}(t, h, x)). \] (C.14)

Plugging (C.12), (C.13) and (C.14) into (C.11), we can see that

\[ \mathbb{P}(\{\Delta X_{u,w_i}(t, h) = k_i\}_{i \in \{1, \ldots, m\}}, |\mathcal{S}| \geq 2 \mid X(t) = x) = o(h). \]

and

\[ \mathbb{P}(\{\Delta X_{u,w_i}(t, h) = k_i\}_{i \in \{1, \ldots, m\}}, |\mathcal{S}| = 1 \mid X(t) = x) = \sum_{i=1}^{m} q_{u,w}(t, x, k_i) h + o(h), \]

where \( \mathcal{S} \) is the set defined in (3.13), \( |\mathcal{S}| \) is the cardinality of \( \mathcal{S} \), and

\[ q_{u,w}(t, x, k_i) = c \frac{\Gamma(x_{u'} + \sum_{i=1}^{m} k_i)}{\Gamma(x_{u'}) \prod_{i=1}^{m} \Gamma(k_i + 1)} \frac{\Gamma(k_i)}{\Gamma(k_i + 1) \Gamma(x_{u'} + \sum_{i=1}^{m} k_i + c)} r_{u,w}(t, x). \]
Appendix D: Proof of Proposition 3.6

(1). Since for each \( i \in \{1, \ldots, m_2\} \), \( \Delta_{u_i'}^{X,u_i}(t, h) \) is independent distributed as (3.16), by Corollary 3.5, when \( c > 2e^{f'_{t+h}r'_{q',q''}(s,x)ds} \), the infinitesimal mean \( \mu_{u_i'}^{X}(t, x) \) is given by

\[
\mu_{u_i'}^{X}(t, x) = x_{u_i'} r_{q',q''}(t, x) \frac{c}{c - 1}
\]

and the infinitesimal variance \( [\sigma_{u_i'}^{X}(t, x)]^2 \) is given by

\[
[\sigma_{u_i'}^{X}(t, x)]^2 = x_{u_i'}^2 r_{q',q''}(t, x) \frac{c}{(c - 1)(c - 2)} + x_{u_i'} r_{q',q''}(t, x) \frac{c}{c - 2}.
\]

Then \( X(t) \) has infinitesimal over-dispersion at \( X(t) = x \) with respect to each arrow in \( (u_i', u_i'') \in \{1, \ldots, m_2\} \). Hence, by Definition 2.3, we know that \( X(t) \) has systemic infinitesimal over-dispersion at \( X(t) = x \) for \( (u_i', u_i'') \in \{1, \ldots, m_2\} \).

Given that \( \Delta_{u_i'}^{X}(t, h) \) and \( \Delta_{u_j'}^{X}(t, h) \) are independent conditional on \( X(t) = x \) and \( \Pi_{q',q''}(t, h, x) \), for \( i, j \in \{1, \ldots, m_2\} \) and \( i \neq j \), we have

\[
\text{Cov}[\Delta_{u_i'}^{X}(t, h), \Delta_{u_j'}^{X}(t, h) \mid X(t) = x, \Pi_{q',q''}(t, h, x)] = 0.
\]

By the law of total covariance, when \( c > 2e^{f'_{t+h}r'_{q',q''}(s,x)ds} \), for \( i, j \in \{1, \ldots, m_2\} \) and \( i \neq j \),

\[
\text{Cov}[\Delta_{u_i'}^{X}(t, h), \Delta_{u_j'}^{X}(t, h) \mid X(t) = x] = \text{Cov}[\mathbb{E}[\Delta_{u_i'}^{X}(t, h) \mid X(t) = x, \Pi_{q',q''}(t, h, x)],
\]

\[
\mathbb{E}[\text{Cov}[\Delta_{u_i'}^{X}(t, h), \Delta_{u_j'}^{X}(t, h) \mid X(t) = x, \Pi_{q',q''}(t, h, x)]
\]

\[
= \text{Cov}[x_{u_i'}(1 - \Pi_{q',q''}(t, h, x))^{-1} \Pi_{q',q''}(t, h, x),
\]

\[
x_{u_j'}(1 - \Pi_{q',q''}(t, h, x))^{-1} \Pi_{q',q''}(t, h, x)] + o(h)
\]

\[
= x_{u_i'} x_{u_j'} \mathbb{E}[1 - \Pi_{q',q''}(t, h, x)]^{-1} \Pi_{q',q''}(t, h, x)]^2
\]

\[
- x_{u_i'} x_{u_j'} \mathbb{E}[1 - \Pi_{q',q''}(t, h, x)]^{-1} \Pi_{q',q''}(t, h, x)]^2 + o(h)
\]

\[
= x_{u_i'} x_{u_j'} \left( \frac{B(\alpha_{q',q''}(t, h, x) + 1, \beta_{q',q''}(t, h, x) - 1)}{B(\alpha_{q',q''}(t, h, x), \beta_{q',q''}(t, h, x))} \right)^2 + o(h).
\]

Note that

\[
\frac{B(\alpha_{q',q''}(t, h, x) + 1, \beta_{q',q''}(t, h, x) - 1)}{B(\alpha_{q',q''}(t, h, x), \beta_{q',q''}(t, h, x))}
\]
Similarly,

$$\frac{B(\alpha_{q^s},\beta_{q^s})(t, h, x) + 2, \beta_{q^s}(t, h, x) - 2)}{B(\alpha_{q^s},\beta_{q^s})(t, h, x)} = \frac{\Gamma(\alpha_{q^s})(t, h, x) + \beta_{q^s}(t, h, x))}{\Gamma(\alpha_{q^s})(t, h, x))\Gamma(\beta_{q^s}(t, h, x))} = \frac{\Gamma(\alpha_{q^s})(t, h, x) + 1)(\beta_{q^s} - 1)}{\Gamma(\alpha_{q^s})(t, h, x))}\Gamma(\beta_{q^s}(t, h, x))$$

Plugging (D.2) and (D.3) into (D.1), we have

$$\text{Cov}[\Delta X_{u_i} (t, h), \Delta X_{u_j} (t, h) | X(t) = x]$$

$$= x_{u_i} x_{u_j} (c(1 - \pi_{q^s}(t, h, x)) - 1)(c(1 - \pi_{q^s}(t, h, x)) - 2)$$

$$- x_{u_i} x_{u_j} \left(\frac{c(1 - \pi_{q^s}(t, h, x)) - 1}{(c-1 - \pi_{q^s}(t, h, x))}\right)^2 + o(h)$$

$$= x_{u_i} x_{u_j} \left(\frac{c(1 - e^{-\int t^t h r_{q^s}(s, x)ds}) + 1)(c(1 - e^{-\int t^t h r_{q^s}(s, x)ds})}{(ce^{-\int t^t h r_{q^s}(s, x)ds} - 1)(ce^{-\int t^t h r_{q^s}(s, x)ds} - 2) - 1}\right)^2 + o(h).$$

By Definition 2.4 and the l’Hôpital’s rule, when $$c > 2e^{\int t^t h r_{q^s}(s, x)ds}$$, for $$i, j \in \{1, \ldots, m_2\}$$ and $$i \neq j$$, the infinitesimal covariance $$\sigma^{X}_{u_i u_j}$$ is given by

$$\sigma^{X}_{u_i u_j}(t, x) = x_{u_i} x_{u_j} c(c - 1)^{-1}(c - 2)^{-1} r_{q^s}(t, x),$$

which is strictly positive.

(2) Since the increments of $$\{N_{u_i}(t)\}_{t \in \{1, \ldots, m_2\}}$$ are distributed as follows:

$$\mathbb{P}(\Delta X_{u_i} (t, h) = k_i | X(t) = x, \Pi_{q^s}(t, h, x))$$
Recall that \( \alpha \) where \( x_{u''} \) hence, we have that for \( P = \Gamma = \{ i \in \{ 1, \ldots, m_2 \} \} \) for all \( i \in \{ 1, \ldots, m_2 \} \), we have

\[
\mathbb{P}(\{ \Delta_{u''}^{X_{u''}}(t, h) = k_i \}_{i \in \{ 1, \ldots, m_2 \}} | X(t) = x) = \int_0^{m_2} \prod_{i=1}^{m_2} \left( x_{u''} + k_i - 1 \right)^{y \sum_{i=1}^{m_2} k_i - \sum_{i=1}^{m_2} x_{u''}} \times \frac{y^{\alpha_{q''}(t,h,x)\alpha_{q''}(t,h,x) \cdot \beta_{q''}(t,h,x) \cdot \beta_{q''}(t,h,x)}}{\Gamma(\alpha_{q''}(t,h,x) \cdot \Gamma(\beta_{q''}(t,h,x)) \cdot \Gamma(\beta_{q''}(t,h,x))} \cdot \frac{\Gamma(\sum_{i=1}^{m_2} k_i + \alpha_{q''}(t,h,x) \cdot \Gamma(\sum_{i=1}^{m_2} x_{u''} + \beta_{q''}(t,h,x) \cdot \beta_{q''}(t,h,x)) \cdot \Gamma(\beta_{q''}(t,h,x) \cdot \beta_{q''}(t,h,x))}{\Gamma(\sum_{i=1}^{m_2} x_{u''} + \sum_{i=1}^{m_2} k_i + c) + o(h)}.
\]

Recall that \( \alpha_{q''}(t,h,x) = c \left( 1 - e^{-\int_{t}^{t + h} r_{q''}(s,x)ds} \right) \), by Taylor series we have

\[
\alpha_{q''}(t,h,x) = (c \cdot r_{q''}(t,x))h + o(h).
\]

Hence, we have that for \( \sum_{i=1}^{m_2} k_i \geq 1 \),

\[
\Gamma \left( \sum_{i=1}^{m_2} k_i + \alpha_{q''}(t,h,x) \right) = \left( \sum_{i=1}^{m_2} k_i + \alpha_{q''}(t,h,x) - 1 \right) \cdot (2 + \alpha_{q''}(t,h,x) - 1) \cdot \alpha_{q''}(t,h,x) = \Gamma(\alpha_{q''}(t,h,x) \cdot \Gamma(\beta_{q''}(t,h,x)) \cdot \Gamma(\beta_{q''}(t,h,x)) \cdot \Gamma(\beta_{q''}(t,h,x)) \cdot \Gamma(\beta_{q''}(t,h,x)) \cdot \Gamma(\beta_{q''}(t,h,x)) + o(\alpha_{q''}(t,h,x)) = \Gamma(\sum_{i=1}^{m_2} k_i) \cdot \alpha_{q''}(t,h,x) \cdot \Gamma(\alpha_{q''}(t,h,x) \cdot \Gamma(\beta_{q''}(t,h,x)) + o(\alpha_{q''}(t,h,x)).
\]

(\text{D.6})
Furthermore, we have
\[
\Gamma\left( \sum_{i=1}^{m_2} x_{u''_i} + \beta_{y''}(t, h, x) \right) \\
= \left( \sum_{i=1}^{m_2} x_{u''_i} - 1 + c - \alpha_{y''}(t, h, x) \right) \cdots \left( c - \alpha_{y''}(t, h, x) \right) \cdot \Gamma(\beta_{y''}(t, h, x)) \\
= \left[ \left( \sum_{i=1}^{m_2} x_{u''_i} - 1 + c \right) \cdots c + O(\alpha_{y''}(t, h, x)) \right] \Gamma(\beta_{y''}(t, h, x)) \\
= \left[ \frac{\Gamma(\sum_{i=1}^{m_2} x_{u''_i} + c)}{\Gamma(c)} + O(\alpha_{y''}(t, h, x)) \right] \Gamma(\beta_{y''}(t, h, x)). \tag{D.7}
\]

Plugging (D.5), (D.6) and (D.7) into (D.4), then when \( \sum_{i=1}^{m_2} x_{u''_i} \geq 1 \)
\[
P(\{ \Delta \mathbf{X}_{u''_i}(t, h) = k_i \}_{i \in \{1, \ldots, m_2\}} | \mathbf{X}(t) = \mathbf{x}) \\
= q(\{ u''_{i_1}, \ldots, u''_{i_{m_2}} \}_{i \in \{1, \ldots, m_2\}}) + o(h),
\]
where the transition rate \( q(\{ u''_{i_1}, \ldots, u''_{i_{m_2}} \}_{i \in \{1, \ldots, m_2\}}) \) is given by
\[
q(\{ u''_{i_1}, \ldots, u''_{i_{m_2}} \}_{i \in \{1, \ldots, m_2\}})(t, x, \{ k_i \}_{i \in \{1, \ldots, m_2\}}) = c \prod_{i=1}^{m_2} \left( x_{u''_i} + k_i - 1 \right)^{k_i} \frac{\Gamma(\sum_{i=1}^{m_2} k_i) \Gamma(\sum_{i=1}^{m_2} x_{u''_i} + c)}{\Gamma(\sum_{i=1}^{m_2} x_{u''_i} + \sum_{i=1}^{m_2} k_i + c) \Gamma(\beta_{y''}(t, h, x))},
\]
which is strictly positive.

Acknowledgments
This research project was supported by NSF grant DMS-1761603. Code and data reproducing our results are available online at https://github.com/patning/Over-dispersion.

References
[1] Anderson, R. M. and May, R. M. (1992). Infectious diseases of humans: Dynamics and control. Oxford university press.
[2] Bjørnstad, O. N., Finkenstädt, B. F., and Grenfell, B. T. (2002). Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. Ecological monographs, 72(2):169–184.
[3] Bjørnstad, O. N. and Grenfell, B. T. (2001). Noisy clockwork: time series analysis of population fluctuations in animals. Science, 293(5530):638–643.
[4] Bretó, C., He, D., Ionides, E. L., and King, A. A. (2009). Time series analysis via mechanistic models. The Annals of Applied Statistics, 3(1):319–348.
[5] Bretó, C. and Ionides, E. L. (2011). Compound Markov counting processes and their applications to modeling infinitesimally over-dispersed systems. *Stochastic Processes and their Applications*, 121(11):2571–2591.

[6] Brown, T. C., Hamza, K., and Xia, A. (1998). On the variance to mean ratio for random variables from Markov chains and point processes. *Journal of applied probability*, pages 303–312.

[7] Gillespie, J. H. (1984). The molecular clock may be an episodic clock. *Proceedings of the National Academy of Sciences*, 81(24):8009–8013.

[8] Guerra, F. M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S. L., Li, Y., and Crowcroft, N. S. (2017). The basic reproduction number (R0) of measles: a systematic review. *The Lancet Infectious Diseases*, 17(12):e420–e428.

[9] He, D., Ionides, E. L., and King, A. A. (2010). Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of the Royal Society Interface*, 7(43):271–283.

[10] Ionides, E. L., Bhadra, A., Atchadé, Y., and King, A. (2011). Iterated filtering. *The Annals of Statistics*, 39(3):1776–1802.

[11] Ionides, E. L., Bretó, C., and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 103(49):18438–18443.

[12] Ionides, E. L., Nguyen, D., Atchadé, Y., Stoev, S., and King, A. A. (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences*, 112(3):719–724.

[13] Kendall, D. G. (1948). On the generalized “birth-and-death” process. *The Annals of Mathematical Statistics*, 19(1):1–15.

[14] King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *Nature*, 454(7206):877–880.

[15] King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69(12).

[16] Koelle, K., Pascual, M., and Yunus, M. (2006). Serotype cycles in cholera dynamics. *Proceedings of the Royal Society B: Biological Sciences*, 273(1603):2879–2886.

[17] Parthasarathy, K. R. (2005). *Probability measures on metric spaces*, volume 352. American Mathematical Soc.
### Table 2
**Notation.**

| Symbol | Description |
|--------|-------------|
| $G = (V, A)$ | $G$: finite graph, $V$: set of vertices, and $A$: set of arrows, Sect. 2.1. |
| $S_o$ | Set of all source vertices, Sect. 2.1. |
| $S_i$ | Set of all sink vertices, Sect. 2.1. |
| $X_v(t)$ | Configuration on vertex $v \in V$ at time $t \in [0, \infty)$, Sect. 2.1. |
| $N_{vv'}^X(t)$ | Nondecreasing integer-valued jump process over $(v, v')$, Sect. 2.1. |
| $\mathcal{C}$ | Finite set of colors, Sect. 2.1. |
| $q(t, x, l)$ | Transition rate function, Sect. 2.2. |
| $q_{(v',v')}(t, x, (k)_i)$ | Transition rate function, Sect. 2.2. |
| $Q(t, x, x')$ | Transition rate function, Sect. 2.2. |
| $\Delta X_{vv'}^X(t, h)$ | Increment of $N_{vv'}^X(t)$ in time interval $[t, t+h]$, Sect. 2.3. |
| $[\sigma^d X_{vv'}^X(t, x)]^2$ | Infinitesimal variance, Def. 2.1. |
| $\mu^d X_{vv'}^X(t, x)$ | Infinitesimal mean, Def. 2.1. |
| $D^d X_{vv'}^X(t, x)$ | Infinitesimal dispersion index, Def. 2.1. |
| $[\sigma X_{vv'}^X(t, x_0)]^2$ | Integrated variance, Def. 2.2. |
| $\mu X_{vv'}^X(t, x_0)$ | Integrated mean, Def. 2.2. |
| $D X_{vv'}^X(t, x_0)$ | Integrated dispersion index, Def. 2.2. |
| $\sigma^d_{uv',vv'}^u(t, x)$ | Infinitesimal covariance over arrows $(u, u')$ and $(v, v')$, Def. 2.4. |
| $\Upsilon_{vv'}(t, z)$ | Transition probability over $(v, v')$, Sect. 3. |
| $r.(t, z)$ | Per-capita rate function. |
| $\tilde{r}.(t, h, z)$ | Transition probability. |
| $\pi.(t, h, z)$ | Transition probability after rewriting. |
| $\Pi.(t, h, x)$ | Stochastic transition probability. |
| $c > 0$ | Inverse noise parameter. |
| $\alpha.(t, h, x)$ | Parameter in the distribution of $\Pi.(t, h, x)$. |
| $\beta.(t, h, x)$ | Parameter in the distribution of $\Pi.(t, h, x)$. |
| $S$ | Set defined in Eqn. (3.5). |
| $\mathcal{F}$ | Set defined in Eqn. (3.13). |
| $\lambda(t, x)$ | Transition rate function, Prop. 3.7. |
| $\mathcal{E}$ | Event: exactly one transition time occurring in $[t, t+h]$, Prop. 3.7. |