UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 11-19
Statistics and Econometrics Series 014
July 2011

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

COMPOUND MARKOV COUNTING PROCESSES AND THEIR APPLICATIONS TO INFINITESIMALLY OVER-DISPERSED SYSTEMS*

Carles Bretó and Edward L. Ionides*****

Abstract

We propose an infinitesimal dispersion index for Markov counting processes. We show that, under standard moment existence conditions, a process is infinitesimally (over-) equi-dispersed if, and only if, it is simple (compound), i.e. it increases in jumps of one (or more) unit(s), even though infinitesimally equi-dispersed processes might be under-, equi- or over-dispersed using previously studied indices. Compound processes arise, for example, when introducing continuous-time white noise to the rates of simple processes resulting in Lévy-driven SDEs. We construct multivariate infinitesimally over dispersed compartment models and queuing networks, suitable for applications where moment constraints inherent to simple processes do not hold.

Keywords: continuous time; counting Markov process; birth-death process; environmental stochasticity; infinitesimal over-dispersion; simultaneous events.

* NOTICE: this is the authors' version after a first round of revision at Stochastic Processes and their Applications; ** Departamento de Estadística and Instituto Flores de Lemus, Universidad Carlos III de Madrid, C/ Madrid 126, 28903 Madrid (Spain), e-mail: carles.breto@uc3m.es; *** Department of Statistics, University of Michigan, 1085 South University Avenue, Ann Arbor, MI 48109-1107, e-mail: ionides@umich.edu.

Compound Markov counting processes and their applications to modeling infinitesimally over-dispersed systems

Carles Bretó[1]

Departamento de Estadística and Instituto Flores de Lemus, Universidad Carlos III de Madrid, C/ Madrid 126, 28903, Madrid, Spain

Edward L. Ionides

Department of Statistics, University of Michigan, 1085 South University Avenue, Ann Arbor, MI 48109-1107

Abstract

We propose an infinitesimal dispersion index for Markov counting processes. We show that, under standard moment existence conditions, a process is infinitesimally (over-) equi-dispersed if, and only if, it is simple (compound), i.e. it increases in jumps of one (or more) unit(s), even though infinitesimally equi-dispersed processes might be under-, equi- or over-dispersed using previously studied indices. Compound processes arise, for example, when introducing continuous-time white noise to the rates of simple processes resulting in Lévy-driven SDEs. We construct multivariate infinitesimally over-dispersed compartment models and queuing networks, suitable for applications where moment constraints inherent to simple processes do not hold.

Keywords: continuous time; counting Markov process; birth-death process; environmental stochasticity; infinitesimal over-dispersion; simultaneous events

1. Introduction

Continuous-time stochastic processes are widely used as a modeling tool for studying dynamical systems in different fields. Most continuous-time processes proposed in the literature belong to one of two large families: real-valued processes which can be written as solutions to stochastic differential equations [24, 30] and discrete-valued processes defined via counting processes [10, 35, 9] or Markov chains [5]. In this paper, we focus on the intersection between counting processes and Markov processes, namely Markov counting processes (MCPs from this point onward). MCPs are building blocks for models which are heavily used in biology (in the context of compartment models) and engineering (in the context of queues and queuing networks) as well as in many other fields.

A *counting process* is a continuous-time, non-decreasing, non-negative, integer-valued stochastic process. The counting process is said to count *events* each of which has an associated *event time*. A counting process is *simple* if, with probability one, there is no time at which two or more events occur simultaneously. A process which is not simple is called *compound*. Simpleness is a convenient, and therefore widely adopted, property for both the theory and applications of counting processes [10]. The Markov property is also a convenient and widespread property of stochastic models. However, we will show that simple MCPs, combining these two attractive properties, have severe limitations in terms of the range of possible relationships between their infinitesimal mean and variance. Previous approaches to negotiate this difficulty have centered on sacrificing the Markov property rather than simpleness. However, there are theoretical and practical attractions to the alternative strategy of maintaining the Markov property while allowing for simultaneous events. Investigating such models is the topic of this paper.

The ratio of the variance to the mean of a random variable is called its *dispersion*. Many well-known integer-valued distributions have dispersion constraints. These constraints are often not reproduced in data from applications,

Email addresses: `carles.breto@uc3m.es` (Carles Bretó), `ionides@umich.edu` (Edward L. Ionides)

[1] corresponding author; Tel. +34916245855; Fax:+34916249849

the data typically having additional variance and therefore being termed *over-dispersed* [29]. The same issues arise in integer-valued stochastic processes [7] and, as a result, there is a considerable literature devoted to extending otherwise appealing models which are unable to reproduce observed variability. Typically, over-dispersion has been studied via defining stochastic processes in which some parameters are themselves modeled as stochastic in order to produce additional variability. This idea has been widely applied since the pioneering work of Greenwood and Yule [19], which derived the over-dispersed negative binomial distribution as a mixture of the Poisson distribution with a gamma-distributed parameter. Another early contribution is the Cox process [8], also known as doubly-stochastic Poisson process [9, 35, 10]. Some recent work has considered stochastic parameters for continuous-time Markov chains [12] and for non-Markovian processes [37]. Marion and Renshaw [28] and Varughese and Fatti [38] studied over-dispersion generated by standard birth-death processes with diffusion-driven rates, focusing on population dynamics applications. Both [28] and [38] proposed a mean-reverting Ornstein-Uhlenbeck process for the driving random environment. Compound counting processes have been studied in the literature on batch processes [31], but we are not aware of a previous investigation of infinitesimal dispersion in this context. To our knowledge, the first general class of infinitesimally over-dispersed MCPs was proposed by Bretó et al. [6]. They achieved over-dispersion by introducing white noise to rates of a multivariate process constructed via simple death processes, which was shown to result in the possibility of simultaneous events. The main goal of this paper is to generalize the model of [6] by presenting a systematic investigation of over-dispersed models via compound MCPs. In particular, we define a class of compound MCPs using Lévy-driven stochastic differential equations [2], by introducing continuous-time white noise in the Kolmogorov equations for simple MCPs. The applications of MCPs are too diverse to cover systematically here. One example, which has been a motivation for our work [6], is the study of infectious disease dynamics. Discrete-state Markov processes have proven useful models for studying many infectious disease transmission systems, and are central to current understanding of the spread of such diseases through populations [25]. However, standard disease models are constructed via simple MCPs and therefore struggle to match the statistical properties observed in data. Recent advances in statistical inference methodology [23, 1] have permitted fitting more general models, based on compound MCPs, to data [6, 21]. At least in this context, the substantial scientific consequences of adequately modeling over-dispersion in stochastic processes are consistent with the widely recognized importance of over-dispersion for drawing correct inferences from integer-valued regression models [29].

As concrete examples of models defined by Lévy-driven Kolmogorov's differential equations, we compute infinitesimal moments and transition rates for various specific novel models. The availability of transition rates makes possible exact simulation, and exact methods are particularly appropriate when dealing with small counts, which arise naturally in some applications. In infectious disease applications, for example, small counts arise at the start of an epidemic, which is a critical period for identifying and controlling the disease transmission. Exact simulation of MCPs can be computationally demanding for processes with a very large number of events. In these cases, it is standard to use approximations which are more affordable computationally but require some diagnostics to investigate the validity of the approximation. To this end, both Euler-Maruyama time discretizations of MCPs and diffusion approximations have been proposed in the literature [6, 21, 23, 26, 28, 38, 13]. Several algorithms have been proposed in which two simulation methods are used, an exact one for small counts and a faster, approximate one for larger counts [20]. In order to use combined algorithms of this type, it is necessary to choose a diffusion approximation, given some MCP. Diffusions are defined in a straightforward way in terms of infinitesimal moments. Requiring that the MCP and the proposed diffusion approximation have common infinitesimal moments gives a natural approach for such an approximation, giving further motivation for the study of infinitesimal moments of MCPs.

A second goal of this paper is to propose the use of an infinitesimal dispersion index for counting processes in conjunction with standard indices. This provides a simple measure of dispersion, combining attractive theoretical properties with scientific interpretability, which is desirable when considering candidate processes for applications. Markov processes specified as the solution to stochastic differential equations are naturally characterized by their infinitesimal mean and variance [24]. However, these infinitesimal moments have not been studied in the context of counting processes, perhaps because, as we will show, in the case of simple MCPs the infinitesimal variance is constrained to be equal to the infinitesimal mean. Instead, interest has focused on dispersion properties of increments of counting processes over fixed time windows, which we call *integrated dispersion* to distinguish it from infinitesimal dispersion. The study of integrally over-dispersed counting processes has a long history, going back at least to the start of the twentieth century [36] and continuing up to the present [e.g., 3]. Integrated dispersion has undoubtedly an interest of its own, in particular if the integration window is chosen according to some specific criterion (possibly motivated in applications by scientific evidence). Because of this window dependence, integrated dispersion may give a distorted representation of a process, in the same way that discretizing a continuous-time process at different resolutions might give very different pictures. In particular, we show that all of integrated over-, equi- and under-dispersion may occur for infinitesimally equi-dispersed processes. By contrast, infinitesimal dispersion provides an intuitive and theoretically attractive measure which has already proven its worth in the study of real-valued Markov processes.

In Section 2 we investigate the infinitesimal moments of simple and compound MCPs, and compare them with previously studied measures of integrated dispersion. Then, in Section 3 we propose several novel over-dispersed compound MCPs. Section 4 demonstrates an application of these MCPs to construct infinitesimally over-dispersed models of disease transmission, motivating the investigation of a general multivariate framework developed further in Appendix D.

2. Dispersion of Markov counting processes

One can study dispersion in the context of non-Markovian processes, but several considerations have led us to focus on the Markov case here. Firstly, there is less room for debate over the definition of appropriate measures of dispersion for Markov processes. Secondly, the extensively studied theory of Markov chains [5] allows us to avoid explicitly discussing measure-theoretic issues while being guaranteed that there are no difficulties concerning the existence and construction of the processes in question. Thirdly, our later goal of studying over-dispersed Markov counting processes clearly does not necessitate a complete investigation of non-Markovian possibilities.

Let $\{N(t) : t \in \mathbb{R}^+\}$ be a time homogeneous Markov counting process, which we will refer to as $\{N(t)\}$. Defining $\Delta N(t) = N(t+h) - N(t)$, the *transition rates* are written as

$$q(n,k) \quad \equiv \quad \lim_{h\downarrow 0} \frac{P(\Delta N(t) = k | N(t) = n)}{h} \tag{1}$$

where $t, h \in \mathbb{R}^+$ and $k, n \in \mathbb{N}$ with $k \geq 1$. The rate of leaving state n or *(infinitesimal) rate function* is written as

$$\lambda(n) \equiv \lim_{h\downarrow 0} \frac{1 - P(\Delta N(t) = 0 | N(t) = n)}{h}.$$

In the Markov Chain literature, the transition rates are also known as the local characteristics of the transition semi-group [5]. In the point process literature, $\lambda(n)$ is the *intensity* of the process and $q(n,k)/\sum_{k\geq 1} q(n,k)$ is the *batch-size distribution* of the simultaneous points [10]. We restrict ourselves to *stable* and *conservative* processes for which $\lambda(n) = \sum_{k\geq 1} q(n,k) < \infty$ for all n. Markov processes satisfying these conditions form a very general class, and the MCP is then characterized by its transition rates [5]. We also restrict ourselves to time homogeneous processes to add clarity to the concepts, results and proofs. However, these can be readily generalized to the non-homogeneous case, for which the transition rates also depend on time.

Measures of dispersion which have previously been considered for counting processes include the variance to mean ratio $V[N(t)]/E[N(t)]$ (for example in [18]) and the difference $V[N(t)] - E[N(t)]$ (in [7]). We will define the integrated dispersion index of $\{N(t)\}$ as

$$D_N(n_0, t) \quad \equiv \quad \frac{V[N(t) - N(0) | N(0) = n_0]}{E[N(t) - N(0) | N(0) = n_0]}. \tag{2}$$

Usually n_0 is assumed to be 0 in which case D_N corresponds to the standard dispersion index defined as a ratio. Note however that (1) defines $\{N(t)\}$ in infinitesimal terms. This suggests the infinitesimal dispersion index which we define as

$$D_{dN}(n) \quad \equiv \quad \frac{\lim_{h\downarrow 0} h^{-1} V[N(t+h) - N(t) | N(t) = n]}{\lim_{h\downarrow 0} h^{-1} E[N(t+h) - N(t) | N(t) = n]} \equiv \frac{\sigma^2_{dN}(n)}{\mu_{dN}(n)}, \tag{3}$$

as an alternative to D_N. The numerator and denominator of (3) are the standard definitions of *infinitesimal variance* and *infinitesimal mean* respectively [24]. We say that $\{N(t)\}$ is infinitesimally equi-dispersed at $N(t) = n$ if $D_{dN}(n) = 1$

and infinitesimally over-dispersed if $D_{dN}(n) > 1$. Correspondingly, we say that $\{N(t)\}$ is integrally equi-dispersed when $D_N(n_0,t) = 1$ and integrally over-dispersed when $D_N(n_0,t) > 1$.

To investigate sufficient conditions for infinitesimal equi-dispersion, we start by considering expressions for the moments of the increments of a process $\{N(t)\}$:

$$\begin{aligned} E[\Delta N^r(t)|N(t)] &= 0^r P(\Delta N(t){=}0|N(t)) + 1^r P(\Delta N(t){=}1|N(t)) + \sum_{k=2}^{\infty} k^r P(\Delta N(t){=}k|N(t)) \\ \lim_{h\downarrow 0} \frac{E[\Delta N^r(t)|N(t)]}{h} &= \lim_{h\downarrow 0} \frac{P(\Delta N(t){=}1|N(t))}{h} + \lim_{h\downarrow 0} \frac{\sum k^r P(\Delta N(t){=}k|N(t))}{h}. \end{aligned} \tag{4}$$

It is straightforward that the difference between any two infinitesimal or integrated moments comes from terms in the sum corresponding to increments of size larger than one, i.e. to simultaneous events. In Theorem 1, we obtain sufficient conditions for infinitesimal equi-dispersion by requiring orderliness in the sense of Daley and Vere-Jones [10, page 47], i.e. $P(\Delta N(t) \geq 2) = o(h)$, and finding circumstances under which the h limit can be exchanged with the limit of the infinite sum in (4). We use the dominated convergence theorem to show that the limits commute under standard moment existence assumptions for a univariate simple MCP, which is therefore equi-dispersed. For example, the Poisson process and counting processes associated with linear birth and death processes are all infinitesimally equi-dispersed. This is true despite the well-known results that the birth and death counting processes are integrally over- and under-dispersed respectively (see Table D1 for more details).

The moment existence conditions we use in our results concern the total number of jumps that a MCP $\{N(t)\}$ makes in an interval $[t, t+\bar{h}]$. Specifically, define a stochastic bound of the infinitesimal rate function $\lambda(N(s))$ by

$$\bar{\Lambda}(t) = \sup_{t\leq s\leq t+\bar{h}} \lambda(N(s)). \tag{5}$$

Here, we suppress the dependence of $\bar{\Lambda}(t)$ on $\bar{h}$. Now consider the following two properties:

P1. For each t and n there is some $\bar{h} > 0$ such that $E[\bar{\Lambda}(t)|N(t) = n] < \infty$.

P2. For each t and n there is some $\bar{h} > 0$ such that $V[\bar{\Lambda}(t)|N(t) = n] < \infty$.

Properties P1 and P2 require that the MCP does not have explosive behavior, and in particular they hold for any *uniform* MCP (i.e., a MCP for which $q(n,k) \equiv q(k)$) in which the jumps are bounded by some k_0 (i.e., when $q(n,k) = 0$ for all $k > k_0$). P1 and P2 also hold for the simple, linear birth process and for the counting process associated to the simple, linear death process (as defined in Table D1).

Theorem 1 (sufficient condition for Markov infinitesimal equi-dispersion)**.** *Let $\{N(t)\}$ be a simple, time homogeneous, stable and conservative Markov counting process. Supposing (P1), the infinitesimal mean is the same as the infinitesimal rate. Supposing (P2), the infinitesimal variance is also the same as the infinitesimal rate, and therefore $\{N(t)\}$ is infinitesimally equi-dispersed.*

A proof of Theorem 1 is given in Appendix A. This theorem demonstrates that simpleness is a sufficient criterion for infinitesimal equi-dispersion of univariate MCPs. The necessity of simpleness for infinitesimal equi-dispersion is presented in Appendix D, together with extensions to multivariate processes. Here, we continue with an extension of Theorem 1 to mixed Markov counting processes, where the transition rates are allowed to depend on a random variable. Such mixtures are a standard way to generate over-dispersion in categorical data models. In the context of point processes, the mixed Poisson process in Daley and Vere-Jones [10] (called a Pólya process by Snyder and Miller [35]) is constructed as a Poisson process with rate M conditional on the mixing random variable M. An immediate result of this mixing is that the resulting process is integrally over-dispersed. It is straightforward to generalize this notion to simple *mixed MCPs*. We construct a mixed simple MCP by first defining a collection $\{N_m(t), m \in \mathbb{R}\}$ of MCPs where $\{N_m(t)\}$ is a simple MCP with rate function $\lambda(n,m)$. A mixing random variable M then defines a mixed simple MCP by setting $\{N(t)\} = \{N_M(t)\}$. We say that $\{N(t)\}$ is stable and conservative if $\{N_m(t)\}$ is stable and conservative for all m. We write $\Lambda(n) \equiv \lambda(n,M)$ for the stochastic rate function of the mixed MCP. It may be surprising that such mixtures of simple MCPs remain infinitesimally equi-dispersed, as we show in Theorem 2. Mixed MCPs are non-Markovian but the measures of dispersion defined in (2) and (3) can still be computed and discussed. For non-Markovian processes, conditioning on the entire past history in (3) could also be considered. For mixed MCPs the conditions P1 and P2 need only the minor modification that $\lambda(n)$ is replaced by $\Lambda(n) = \lambda(n, M)$. Specifically, we define the stochastic bound $\bar{\Lambda}^*(t) = \sup_{t \leq s \leq t+\bar{h}} \Lambda(N(s))$ and set

P1*. For each t and n there is some $\bar{h} > 0$ such that $E[\bar{\Lambda}^*(t)|N(t) = n] < \infty$.

P2*. For each t and n there is some $\bar{h} > 0$ such that $V[\bar{\Lambda}^*(t)|N(t) = n] < \infty$.

In the context of mixed MCPs, P1* or P2* imply that the mean rate function is finite for each n, $E[\Lambda(n)] < \infty$, so the tails of the additional randomness resulting from M are required to be not too heavy. A proof of Theorem 2 follows a similar approach to Theorem 1 and is available in Appendix D.

Theorem 2 (sufficient condition for mixed Markov infinitesimal equi-dispersion)**.** *Let $\{N(t)\}$ be a time homogeneous, stable and conservative mixed simple MCP, constructed as above with a mixing random variable M. Supposing (P1*), the infinitesimal mean is the same as the average infinitesimal rate. Supposing (P2*), the infinitesimal variance is also the same as the average infinitesimal rate, and therefore $\{N(t)\}$ is infinitesimally equi-dispersed.*

3. Over-dispersed Markov counting processes

From Section 2, we know that simple MCPs are infinitesimally equi-dispersed under standard moment conditions. We therefore seek to generalize standard simple MCP models, to relax this dispersion constraint. Our first approach is to investigate random time change, or subordination, which we show can be interpreted as the inclusion of continuous-time noise in the rate function. Section 3.1 uses this technique to construct some specific processes. In Section 3.2 we consider a subtly different approach of defining an over-dispersed MCP via the limit of a sequence of processes in which discrete-time noise is used to modify the rate.

We know from Section 2 that introducing noise via a mixing random variable in the rate function does not alter the infinitesimal equi-dispersion of simple MCPs. This suggests considering more complex, alternative noise processes. One possibility is to introduce some continuous-time process, say $\{\eta(t)\}$, in the rate function of the MCP. Such constructions may be expected to give processes which are Markov conditional on $\{\eta(t)\}$ but not unconditionally. Our approach is similar to that of [28] and [38]; we propose defining a process by replacing $\lambda(n)$, the deterministic rate function of the original MCP, in Kolmogorov's backward differential system by the stochastic process $\{\lambda(n, \eta(t))\}$ (see Appendix B for a formal definition). However, by taking $\{\eta(t)\}$ to be a suitable white noise process, we differ from [28] and [38] by constructing processes which will be shown to be unconditionally Markov. The consideration of non-white noise is no doubt appropriate in some applications, but white noise provides a relatively simple extension to infinitesimally equi-dispersed processes controlled by a single parameter for the magnitude of the noise. Staying within the class of Markov processes also facilities both theoretical and numerical analysis of the resulting models.

The noise process $\{\eta(t)\}$ could enter $\lambda(n)$ additively or multiplicatively. Given the non-negativity constraint on the infinitesimal rate functions, multiplicative non-negative noise is a simple and convenient choice. We refer to white noise, $\{\xi(t)\} \equiv \{dL(t)/dt\}$, as the derivative of an *integrated noise* process $\{L(t)\}$ which has stationary independent increments. Note that we do not necessarily require that the mean of $L(t)$ is zero. Although $\{\xi(t)\}$ may not exist, in the sense that $\{L(t)\}$ may not have differentiable sample paths, $\{\xi(t)\}$ can nevertheless be given formal meaning [24, 6]. Restricting $\{\xi(t)\}$ to non-negative white noise, the family of increasing Lévy processes provides a rich class from which to choose the integrated noise $\{L(t)\}$. Multiplicative unbiased noise is achieved by requiring $E[L(t)] = t$, in which case$\lim_{h \downarrow 0} E[\Delta L(t)]/h = 1$.

From an alternative perspective, in the context of the general theory of Markov processes, random time change or subordination of an initial process is a well established tool to obtain new processes. Following Sato [33], let $\{M(t)\}$ (the *directing* process) be a temporally homogeneous Markov process and $\{L(t)\}$ (the *subordinator*) be an increasing Lévy process. Any temporally homogeneous Markov process $\{N(t)\}$ identical in law to $\{M \circ L(t)\} \equiv \{M(L(t))\}$ is said to be *subordinate* to $\{M(t)\}$ by the subordinator $\{L(t)\}$.

Theorem 3 below formally states that subordinate processes to simple (and hence equi-dispersed) MCPs are equivalent to solutions of Lévy-driven stochastic differential equations resulting from introducing unbiased multiplicative Lévy white-noise in the deterministic Kolmogorov backward differential system of the directing process. This gives us a license to interpret noise on the rate of a MCP as subordination of the MCP to a Lévy process. This guarantees, under the very general condition of continuity of the transition probabilities of the directing process $\{M(t)\}$, that the resulting subordinate process will remain Markovian [14, page 347]. In Subsection 3.1, we obtain exact results when investigating concrete examples of over-dispersion by exploiting this connection between gamma white noise in the rates and gamma subordinators. To preserve the flow of the main themes of this paper, the technical details involved in the link between subordination and stochastic rates, including definitions of the quantities in Theorem 3 and the proof, are deferred to Appendix B.

Theorem 3 (Lévy white noise and subordination)**.** *Consider the simple, time homogeneous, stable and conservative Markov counting process $\{M_\lambda(t)\}$ defined by the rate function $\lambda(m)$. Let $\{L(t)\}$ be a non-decreasing, Lévy process with $L(0) = 0$ and $E[L(t)] = t$. Let $\{M_{\lambda\xi}(t)\}$ be the process resulting from introducing unbiased, non-negative, multiplicative, Lévy white-noise $\{\xi(t)\} \equiv \{dL(t)/dt\}$ in the rate of $\{M_\lambda(t)\}$, defined as the solution to the Lévy-driven Kolmogorov backward differential system in (B.3). Then, if this solution exists and is unique,*

$$M_{\lambda\xi}(t) \sim M_\lambda \circ L(t) \sim M_\lambda\Big(\int_0^t \xi(u)\, du\Big).$$

3.1. Subordinate processes to simple MCPs by gamma subordinators

A convenient candidate for non-negative continuous-time noise is gamma noise. In this case, the integrated noise process $L(t) = \Gamma(t)$ is a Gamma process defined to have independent, stationary increments with $\Gamma(t) - \Gamma(s) \sim \mathrm{Gamma}\,([t-s]/\tau, \tau)$. Here, $\mathrm{Gamma}\,(a, b)$ is the gamma distribution with mean ab and variance ab^2. In Subsections 3.1.1–3.1.3, we study the inclusion of gamma noise in the rates of three widely used infinitesimally equi-dispersed processes: The Poisson process, linear birth process and linear death process.

Following the convention for naming of subordinate processes [33], we will place the name of the original process first, followed by the name of the driving subordinating noise. We have chosen to study in detail the Poisson, linear birth and linear death processes because they are basic blocks widely used to build more complex, multi-process models, such as compartmental models used in population dynamics and queuing networks in engineering. What makes these three processes fundamental is that they capture in the simplest way, i.e. linearly, the most common possibilities in real applications. Namely, events that by occurring “kill” the potential for future events (death process, or negative feedback); events that “reproduce” meaning that their occurrence fuels that of future events (birth process, or positive feedback); and events which occur independently of the events which have already happened (Poisson, or immigration process, or no feedback). For these processes we provide three results: their first two moments about the mean, which show they are indeed infinitesimally over-dispersed; the distribution of the counting process, which allows for exact, direct simulation of the counting process; and a closed form for the transition rates, which fully characterize the processes and may be used for exact simulation of the event times of the point process and for indirect, exact simulation of the counting process by aggregation.

3.1.1. The Poisson gamma process

We construct an infinitesimally over-dispersed Poisson process. This is a special case of the general compound Poisson process [10], which can be constructed as independent jumps from an arbitrary distribution occurring at the times of a Poisson process. Our alternative construction, derived through introducing white noise on the rate, has an advantage that it can be applied (as we show) not just to Poisson processes but to more general univariate and multivariate processes.

Proposition 4 (Poisson gamma process)**.** *Let $\{M(t)\}$ be a MCP with $q(m,1) = \alpha$ and $q(m,k) = 0$ for $k > 1$, i.e. a time homogeneous Poisson process with rate α. Let $\{\xi(t)\} \equiv \{d\Gamma(t)/dt\}$ be continuous-time gamma noise, where $\Gamma(t) \sim$ Gamma $(t/\tau, \tau)$ with τ parameterizing the magnitude of the noise. Define the subordinate process $\{N(t)\} = \{M(\Gamma(t))\}$, which can be interpreted via Theorem 3 as a Poisson process with a stochastic rate $\alpha\xi(t)$. $\{N(t)\}$ is a compound infinitesimally over-dispersed MCP with increment probabilities*

$$P(\Delta N(t) = k | N(t) = n) \;=\; \frac{G(\tau^{-1}h + k)}{k!\, G(\tau^{-1}h)} p^{\tau^{-1}h} (1-p)^k, \quad \text{for } k \in \mathbb{N},$$

where $p = (1+\tau\alpha)^{-1}$ and G is the gamma function. The transition rates are $q(n,k) = (\tau k)^{-1}(1-p)^k$ for $k \geq 1$. The intensity is $\lambda(n) = \tau^{-1}\log(p^{-1})$. The infinitesimal moments are $\mu_{dN}(n) = \alpha$ and $\sigma^2_{dN}(n) = \alpha(1+\tau\alpha)$ with infinitesimal dispersion $D_{dN}(n) = 1 + \tau\alpha$.

3.1.2. The binomial gamma process

Here, we consider multiplicative gamma noise on the rate of a linear death process. This process has been proposed as a model for biological populations [6], although it was defined as the limit of discrete-time stochastic processes rather than as the solution to the Lévy-driven Kolmogorov differential system of (B.3). It is standard to define death processes as decreasing processes, however our general framework constructs counting processes which are necessarily increasing. Thus, we count cumulative deaths rather than recording the size of the remaining population.

Proposition 5 (binomial gamma process)**.** *Let $\{M(t)\}$ be a MCP with $q(m,1) = \delta(d_0 - m)\,\mathbb{I}\{m < d_0\}$ and $q(m,k) = 0$ for $k > 1$, i.e. the counting process associated with a linear death process having individual death rate $\delta \in R^+$ and initial population size $d_0 \in \mathbb{N}$. Let $\{\xi(t)\} \equiv \{d\Gamma(t)/dt\}$ be continuous-time gamma noise, where $\Gamma(t) \sim Gamma\,(t/\tau, \tau)$ with τ parameterizing the magnitude of the noise. Define $\{N(t)\} = \{M(\Gamma(t))\}$, which corresponds to a death process with a stochastic rate $\delta\xi(t)$ via Theorem 3. $\{N(t)\}$ is a compound infinitesimally over-dispersed MCP with increment probabilities*

$$P(\Delta N(t) = k | N(t) = n) \quad = \quad \binom{d_0 - n}{k} \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} \Big(1 + \delta\tau(d_0 - n - j)\Big)^{-h\tau^{-1}}, \quad \text{for } k \in \{0, \dots, d_0 - n\},$$

and transition rates

$$q(n,k) \quad = \quad \binom{d_0 - n}{k} \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j+1} \tau^{-1} \ln\left(1 + \delta\tau(d_0 - n - j)\right), \quad \text{for } k \in \{1, \dots, d_0 - n\}, \tag{6}$$

for $n < d_0$. The intensity is $\lambda(n) = \tau^{-1}\ln\left(1 + \delta\tau(d_0 - n)\right)$ and the infinitesimal moments and dispersion are

$$\mu_{dN}(n) = (d_0 - n)\tau^{-1}\ln(1 + \delta\tau), \qquad \sigma^2_{dN}(n) = \mu_{dN}(n) + (d_0 - n)\tau^{-1}\left[(d_0 - n - 1)\ln\left(\frac{(1 + \delta\tau)^2}{1 + 2\delta\tau}\right)\right],$$

$$D_{dN}(n) = 1 + (d_0 - n - 1)\left[\frac{2\ln(1 + \delta\tau) - \ln(1 + 2\delta\tau)}{\ln(1 + \delta\tau)}\right].$$

Hence, $\{N(t)\}$ is infinitesimally over-dispersed for $N(t) < d_0 - 1$ and equi-dispersed for $N(t) = d_0 - 1$.

We see from Proposition 5 that the binomial gamma process is infinitesimally over-dispersed as long as there is more than one individual alive from the initial population of d_0 individuals. However, this over-dispersion decreases with the number of remaining individuals, until reaching infinitesimal equi-dispersion when there is only one individual left.

3.1.3. The negative binomial gamma process

Unlike for the death process, when introducing gamma noise to the birth process we are only able to show existence of moments imposing a restriction on the parameter space. In particular, the birth rate of the original process imposes an upper bound on the infinitesimal over-dispersion. When this restriction does not hold, the moments of the resulting process do not exist, and hence our dispersion index is not defined. We include the derivations with gamma noise for consistency with the over-dispersed Poisson and death process. Considering a common subordinator for all three processes has the advantage that it leads naturally to multivariate extensions in which over-dispersed univariate processes are combined to construct multivariate models. It would be possible to use other subordinators, such as the inverse Gaussian process, for which the moment generating function is available in closed form.

Proposition 6 (negative binomial gamma process)**.** *Let $\{M(t)\}$ be a MCP with $q(m,1) = \beta m\,\mathbb{I}\{m > 0\}$ and $q(m,k) = 0$ for $k > 1$, i.e. a linear birth process having individual birth rate $\beta \in \mathbb{R}^+$. Let $\{\xi(t)\} \equiv \{d\Gamma(t)/dt\}$ be continuous-time gamma noise, where $\Gamma(t) \sim Gamma\,(t/\tau, \tau)$ with τ parameterizing the magnitude of the noise. Define $\{N(t)\} = \{M(\Gamma(t))\}$, which corresponds to a birth process with a stochastic rate $\beta\xi(t)$ via Theorem 3. $\{N(t)\}$ is a compound infinitesimally over-dispersed MCP with increment probabilities*

$$P(\Delta N(t) = k | N(t) = n) \quad = \quad \binom{n + k - 1}{k} \sum_{j=0}^{k} \binom{k}{j} (-1)^{k-j} \Big(1 + \beta\tau(n + k - j)\Big)^{-h\tau^{-1}}, \quad \text{for } k \in \mathbb{N},$$

and transition rates

$$q(n,k) \;=\; \binom{n+k-1}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j+1}\tau^{-1}\ln\left(1+\beta\tau(n+k-j)\right), \quad \textit{for } k \geq 1,$$

for $n \in \mathbb{N}$. *The intensity is* $\lambda(n) = \tau^{-1}\ln(1+\beta\tau n)$. *For* $2\beta\tau < 1$, *the infinitesimal moments and dispersion are*

$$\mu_{dN}(n) = n\tau^{-1}\ln\left(\frac{1}{1-\beta\tau}\right), \qquad \sigma^2_{dN}(n) = \mu_{dN} + n\tau^{-1}\left[(n-1)\ln\left(\frac{(1-\beta\tau)^2}{1-2\beta\tau}\right)\right],$$

$$D_{dN}(n) = 1 + (n-1)\left[\frac{2\ln(1-\beta\tau) - \ln(1-2\beta\tau)}{-\ln(1-\beta\tau)}\right].$$

Hence, $\{N(t)\}$ *is infinitesimally over-dispersed for* $N(t) > 1$ *and equi-dispersed for* $N(t) = 1$.

3.1.4. The bivariate binomial gamma process

In some special cases, subordination can enable construction of multivariate over-dispersed MCPs. A simple example of this occurs when a single subordinator is applied to two independent simple MCPs. This corresponds to two processes which are conditionally independent but have the same noise process applied to each rate. Such processes arise, for example, in disease transmission models [26]: if new infections are split into mild and severe cases, then both processes will suffer the same transition rate variability due to weather and heterogeneous human social aggregation. In Proposition 7, we consider the case of two death processes coupled by sharing a common stochastic per-capita death rate. More generally, one could consider dependent processes with different transition rates and affected by non-identical but correlated noise processes. The proof of Proposition 7 is similar to the proof of Proposition 5, and is available in Appendix D. In the statement and proof of the proposition, we implicitly extend the definition given for transition rate in (1) to the bivariate case. This extension is formalized later in equation (11).

Proposition 7 (bivariate binomial gamma process)**.** *Let* $\{M_1(t)\}$ *and* $\{M_2(t)\}$ *be independent MCPs with* $q(m_i,1) = \delta(d_{0i}-m_i)\,\mathbb{I}\{m_i < d_{0i}\}$ *and* $q(m_i,k_i) = 0$ *for* $k_i > 1$ *and* $i = 1,2$, *i.e. the counting processes associated with two independent linear death processes having an equal individual death rate* $\delta \in R^+$ *and initial population sizes* $d_{0i} \in \mathbb{N}$. *Let* $\{\xi(t)\} \equiv \{d\Gamma(t)/dt\}$ *be continuous-time gamma noise, where* $\Gamma(t) \sim Gamma\,(t/\tau,\tau)$ *with* τ *parameterizing the magnitude of the noise. Define* $\{N_i(t)\} = \{M_i(\Gamma(t))\}$, *for* $i = 1,2$, *corresponding to two death processes each having stochastic rate* $\delta\xi(t)$ *via Theorem 3.* $\{N_1(t)\}$ *and* $\{N_2(t)\}$ *are compound infinitesimally over-dispersed MCPs with joint increment probabilities*

$$P(\Delta N_i(t) = k_i | N_i(t) = n_i) \;=\; \binom{d_{01}-n_1}{k_1}\binom{d_{02}-n_2}{k_2}\sum_{j=0}^{k_1+k_2}\binom{k_1+k_2}{j}(-1)^{k_1+k_2-j}\Big(1+\delta\tau(d_{01}+d_{02}-n_1-n_2-j)\Big)^{-h\tau^{-1}}$$

for $k_i \in \{0,\ldots,d_{0i}-n_i\}$. *The transition rates of the bivariate Markov chain* $\{N_1(t), N_2(t)\}$ *are*

$$q((n_1,n_2),(k_1,k_2)) \;=\; \binom{d_{01}-n_1}{k_1}\binom{d_{02}-n_2}{k_2}\sum_{j=0}^{k_1+k_2}\binom{k_1+k_2}{j}(-1)^{k_1+k_2-j+1}\tau^{-1}\ln\left(1+\delta\tau(d_{01}+d_{02}-n_1-n_2-j)\right)$$

for $n_i < d_{0i}$ *and* $k_i \in \{1,\ldots,d_{0i}-n_i\}$. *The marginal transition rates, increment probabilities and infinitesimal moments of* $\{N_i(t)\}$ *are the same as those of a binomial gamma process. The infinitesimal covariance of* $\{N_1(t)\}$ *and* $\{N_2(t)\}$ *is*

$$\begin{aligned}\sigma_{dN_1dN_2}(n_1,n_2) &= \lim_{h\downarrow 0} h^{-1}Cov[N_1(t+h)-N_1(t), N_2(t+h)-N_2(t) | N_1(t)=n_1, N_2(t)=n_2] \\ &= (d_{01}-n_1)(d_{02}-n_2)\tau^{-1}\ln\left(\frac{(1+\delta\tau)^2}{1+2\delta\tau}\right) \;>\; 0.\end{aligned}$$

3.2. The binomial beta process

An alternative to the subordination construction of Section 3.1 is to define noise on the rate of a continuous-time MCP via a limit of discrete-time processes [6]. Here, we investigate a specific example of an infinitesimally over-dispersed death process. As we will see in Section 4, infinitesimally over-dispersed death processes play a useful role in constructing over-dispersed multivariate models. An inconvenience of the binomial gamma process is that its infinitesimal moments are a non-linear system of two equations which, to obtain a desired mean and variance, needs to be solved numerically. A moment-based parameterization allows to easily change the variability (via the variance) for a fixed location (fixing the mean), facilitating the interpretation of the parameters. In the context of counting processes, such parameterization has the additional advantage that it permits a direct and straightforward comparison with analogous stochastic differential equations. This motivates an alternative over-dispersed death process model, the binomial beta process defined below, which can be easily parameterized in terms of the infinitesimal moments.

Instead of introducing continuous-time noise to the rates, we consider introducing it directly to the transition probabilities of the death process. Since probabilities must be constrained to the unit interval we need to consider an alternative to gamma noise, and a convenient choice is beta noise. The construction of a beta process as a process with beta independent increments is not, however, straightforward [22]. Our construction of compound processes based on noise introduction consisting on taking limits of discrete-time processes avoids this difficulty. Let $\{M(t)\}$ be a MCP with $q(m, 1) = \delta(d_0 - m)\,\mathbb{I}\{m < d_0\}$ and $q(m, k) = 0$ for $k > 1$, i.e. the counting process associated with a linear death process having individual death rate $\delta \in R^+$ and initial population size $d_0 \in \mathbb{N}$. The increments of $\{M(t)\}$ are binomially distributed,

$$M(t + h) - M(t)|M(t) \sim \text{Binom}(d_0 - M(t), \pi(h)) \qquad \text{for } \pi(h) = 1 - e^{-\delta h}.$$

We consider a discrete-time process $\{M_h(t), t \in \{0, h, 2h, \dots\}\}$, defined by constructing a sequence of independent, identically distributed random variables $\Pi_0, \Pi_1, \dots$ and setting

$$M_h((i + 1)h) - M_h(ih) \,\big|\, M_h(ih), \Pi_i \sim \text{Binom}(d_0 - M_h(ih), \Pi_i).$$

To interpret Π_i as noise added to $\pi(h)$, we require $E[\Pi_i] = \pi(h)$ and $\Pi_i \in [0, 1]$. If a finite limit

$$q(n, k) = \lim_{h \to 0} h^{-1}\Big[P(M_h(h) - M_h(0) = k|M_h(0) = n)\Big]$$

exists for each $n < d_0$ and $k \in \{1, \dots, d_0 - n\}$, with $\sum_k q(n, k) < \infty$, then we define $\{N(t)\}$ to be a MCP with these transition rates. The binomial beta process corresponds to the specific choice $\Pi_i \sim \text{Beta}(a, b)$, the beta distribution having mean $a/(a+b)$ and variance $ab/\big((a+b)^2(a+b+1)\big)$, with $a = c\pi(h)$ and $b = c(1-\pi(h))$. Here, $c > 0$ is an inverse noise parameter in the sense that it does not affect the mean of Π_i and the variance $V[\Pi_i] = \pi(h)(1 - \pi(h))(c + 1)^{-1}$ is a decreasing function of c. Proposition 8 shows that this choice leads to a well-defined process and identifies its infinitesimal mean and variance.

Proposition 8 (binomial beta process)**.** *Let $\{N(t)\}$ be the binomial beta process constructed above, with initial population size d_0, individual death rate δ and noise parameter c. $\{N(t)\}$ is an infinitesimally over-dispersed compound Markov counting process with transition rates being*

$$q(n, k) \quad = \quad \binom{d_0 - n}{k} \frac{G(k)\, G(c + d_0 - n - k)}{G(c + d_0 - n)} c\delta$$

for $n < d_0$ and $k \in \{1, \dots, d_0 - n\}$ and zero otherwise. The infinitesimal mean is $\mu_{dN}(n) = (d_0 - n)\delta$. The infinitesimal variance is $\sigma^2_{dN}(n) = (1 + \omega(n))\mu_{dN}(n)$ where $\omega(n) = (d_0 - n - 1)(c + 1)^{-1}$. The infinitesimal dispersion is therefore $D_{dN}(n) = 1 + \omega(n)$, and so the process is infinitesimally over-dispersed when $N(t) < d_0 - 1$ and, since $\omega(d_0 - 1) = 0$, it is equi-dispersed when $N(t) = d_0 - 1$.

Although the infinitesimal moments of the binomial beta process have a simpler form than those of the binomial gamma process, we no longer obtain an expression for the increment probabilities. Hence, exact simulation of the counts is only possible (with the present results) by aggregation from exact event time simulation based on the provided transition rates.

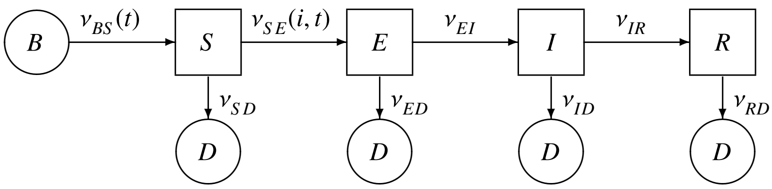

Figure 1: Flow diagram for measles. The population is divided into four compartments $\{S, E, I, R\}$ corresponding to susceptible, exposed, infected and recovered individuals respectively. Two auxiliary nodes $\{B, D\}$ are used to represent the birth and death of individuals. Arrows represent possible transitions, and their labels parameterize the corresponding transition rates.

A similar construction of a negative binomial beta process would follow by adding noise to the individual birth event probability for a sequence of discrete-time birth process, i.e. defining a sequence of random variables with mean $\pi(h) = 1 - e^{-\beta h}$ taking values in the unit interval. As for the binomial beta process, this construction would provide a more convenient parameterization in terms of infinitesimal moments.

4. Markov counting systems and their infinitesimal dispersion

Models for queues, networks of queues, or biological systems may involve multiple interacting counting processes. We discuss such processes in the context of an example drawn from the study of infectious diseases, specifically measles transmission dynamics. This example is presented as a scientific application of the processes developed in Section 3. However, it also motivates a general model for dependent MCPs which we call a Markov counting system. We propose a definition of infinitesimal dispersion for such a system which extends the definition for a MCP. One can go beyond this to extend the formalities of Sections 2 and 3 to these systems, and such a program is described in Appendix D.

There are two distinct motivations for modeling simultaneous events. Firstly, the process in question may indeed have such occurrences, and such applications arise in modeling production systems and rental businesses [31], physical processes and quantum optics [16] and internet traffic [27]. Secondly, the process may have clusters of event times that are short compared to the scale of primary interest. For example, in some applications only aggregated counts and not event times are available, and in this case any clustering time scale which is shorter than the aggregation timescale may be appropriately modeled by simultaneous events. Modeling disease transmission falls into this latter category. It is hard to imagine having exact event time data, and at sufficiently fine time scales the events of becoming infected, infectious and recovered are not even well defined. However, it is easy to imagine clusters of event times, for example, multiple infections caused by a sneeze on a crowded bus. The conceptual, theoretical and computational convenience of Markov processes has led to their widespread use for modeling disease transmission processes. In the past, a modeling hypothesis that events occur non-simultaneously has been favored. This assumption, which has often been made without much consideration of alternatives, can eliminate the possibility of a good statistical fit to data, since we have shown that simultaneous events are required in order for MCPs and Markov counting systems to obtain the over-dispersion observed in data.

Worldwide, measles remains a leading cause of vaccine-preventable death and disability [15]. Global eradication of this highly infectious disease, by intensive vaccination, would be difficult but perhaps not impossible [11]. A fundamental class of models for measles transmission places each individual into exactly one of four compartments termed susceptible (S), exposed and carrying a latent infection (E), infected and infectious (I) and recovered or removed (R). Much previous analysis of SEIR models has employed continuous-time Markov chain models in which simultaneous transitions are assumed not to occur. Based on the results of Section 2, one might expect that such models rule out infinitesimal over-dispersion in a multivariate sense. SEIR-type Markov chain models for disease transmission with gamma noise added to the transition rates have been demonstrated to give improved fit to data [6, 21]. Here, we show how these stochastic rate Markov chain models can be constructed and generalized using infinitesimally over-dispersed MCPs.

The weighted directed graph in Figure 1 gives a diagrammatic representation of a SEIR model. The four compartments $\{S, E, I, R\}$ are represented by boxed nodes. The circled nodes $\{B, D\}$ are used notationally to discuss flows into and out of the population, representing the biological birth and death of individuals. Arrows are used to indicate the possibility of transitions between nodes, with labels parameterizing the transition rates. The state of the system at time t is given by the number of individuals in each compartment and is written as $(S(t), E(t), I(t), R(t))$. The standard interpretation of Figure 1 as a Markov chain [25] has transition rates conditional on $\boldsymbol{X}(t) = (s, e, i, r)$, using the notation of equation (11) below, given by

$$Q[(s,e,i,r),(s-1,e+1,i,r)] = \nu_{SE}(i,t)\, s\, \mathbb{I}\{s>0\}, \tag{7}$$

$$Q[(s,e,i,r),(s+1,e,i,r)] = \nu_{BS}(t), \tag{8}$$

$$Q[(s,e,i,r),(s,e-1,i+1,r)] = \nu_{EI}\, e\, \mathbb{I}\{e>0\}, \qquad Q[(s,e,i,r),(s,e,i-1,r+1)] = \nu_{IR}\, i\, \mathbb{I}\{i>0\},$$

$$q[(s,e,i,r),(s-1,e,i,r)] = \nu_{SD}\, s\, \mathbb{I}\{s>0\}, \qquad Q[(s,e,i,r),(s,e-1,i,r)] = \nu_{ED}\, e\, \mathbb{I}\{e>0\},$$

$$Q[(s,e,i,r),(s,e,i-1,r)] = \nu_{ID}\, i\, \mathbb{I}\{i>0\}, \qquad Q[(s,e,i,r),(s,e,i,r-1)] = \nu_{RD}\, r\, \mathbb{I}\{r>0\},$$

and zero otherwise. Here, $\nu_{SE}(i,t) = \beta(t)(i+\gamma)$ where $\beta(t)$ is a seasonally-varying constant and γ models an external source of infections. The population birth rate $\nu_{BS}(t)$ and per-capita death rate $\nu_{SD} = \nu_{ED} = \nu_{ID} = \nu_{RD}$ are considered known from census data. Mortality from measles has become negligible in developed countries and is ignored. The remaining per-capita rate parameters, ν_{EI} and ν_{IR}, are treated as constants. Bretó et al. [6] and He et al. [21] proposed introducing gamma noise to $\nu_{SE}(i,t)$ by replacing (7) with

$$Q[(s,e,i,r),(s-k,e+k,i,r)] = \binom{s}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j+1}\tau^{-1}\ln\left(1+\nu_{SE}(i,t)\tau(s-j)\right) \quad \text{for } k=1,\ldots,s. \tag{9}$$

This may be interpreted as the transition rate of the binomial gamma process in equation (6) with remaining population size $d_0 - n = s$ and with per-capita 'death' rate $\delta = \nu_{SE}(i,t)$. Here, 'deaths' correspond to transitions from S to E, and at time t there are s individuals available to make this transition. The magnitude of the noise on the transition rate is parameterized by τ, and (9) reduces to (7) in the limit as $\tau \to 0$.This interpretation of (9) provides a general approach for introducing white noise to rates that depend on time and the state of other compartments. One can first add white noise to an appropriate homogeneous process, as carried out in Section 3. Then one can introduce dependencies by making the fixed parameters of the homogeneous process functions of time (e.g., in $\nu_{BS}(t)$), the current state of the system or both (e.g., in $\nu_{SE}(i,t)$). We thereby avoid the consideration of subordination in multivariate and time-inhomogeneous settings. Replacing (7) by (9) only affects the infinitesimal properties (mean, variance and dispersion) of the SE transitions but not those of other transitions. This assertion is checked as part of the formal treatment of the multivariate extension in Appendix D. As a consequence, the interpretation of the rest of transition rates and parameters remains unaltered by the addition of the infinitesimal variability. Also, the infinitesimal properties of the SE transition follow directly from the univariate results (already available from Section 3), unaffected by the other transitions or by whether independent additional variability is added to them.

Inference for either the standard model or the over-dispersed version (9) is complicated by the availability of only incomplete observations consisting of biweekly case reports. Modeling these data as the newly infected individuals in the corresponding time period, scaled by a reporting rate and perturbed by a measurement error, gives rise to a partially observed Markov process known as a state space model [34]. The likelihood of such a nonlinear state space model may be evaluated and maximized by Monte Carlo methods. He et al [21] carried out such a procedure for measles data from London, England, and obtained a 95% confidence interval for τ of $[0.053, 0.10]$. Including over-dispersion is therefore mandated for the data, in the context of this model. A specific consequence of failure to include over-dispersion is that parameter estimates are biased toward models that allow increased stochastic variability. In our measles example, the mean infectious and latent periods were estimated to be close to the lower range of previously available estimates based on clinical experiments when over-dispersion was not allowed for. When allowing for over-dispersion, the estimates changed to the upper range [21]. The effect of over-dispersion on parameter estimates is a consequence of nonlinearity in the model. For linear models, over-dispersion affects estimates of the uncertainty of parameter estimates but not the estimates themselves [29].

If one wished to add noise to the population birth rate $\nu_{BS}(t)$, one could modify the inhomogeneous Poisson process specified in (8) by using the transition rates for the Poisson gamma process of Section 3. If births were modeled as a per-capita rate, rather than based on census data, then the natural simple model is $Q[(s,e,i,r),(s+1,e,i,r)] =$

$\nu_{BS}(t)p\,\mathbb{I}\{p > 0\}$ where $p = s + e + i + r$. To include noise in the rate of this negative binomial process (i.e., linear birth process), one could make use of the negative binomial gamma process. The binomial beta process provides an alternative construction for modeling stochastic rates in (7).

From a biological perspective, mechanisms for over-dispersion in the force of infection could be variations in transmissibility (modulated by temperature or humidity) or variations in contact rates between individuals (due to aggregations such as large sporting events). From a statistical perspective, any misspecification could result in the data requesting over-dispersion when fitting a model. This is an instance of the common statistical issue of distinguishing the extent to which model prediction errors are driven by model misspecification (i.e., imperfect knowledge of the system) versus genuine stochasticity in the system.

To generalize this measles transmission example, let C be a finite collection of compartments, corresponding to $C = \{S, E, I, R\}$ in Figure 1. Let $\mathcal{A}$ be a finite collection of auxiliary nodes, corresponding to $\mathcal{A} = \{B, D\}$ in Figure 1. The collection of possible transitions consists of pairs of compartments/nodes in $\mathcal{T} \subset (C \times C) \cup (C \times \mathcal{A}) \cup (\mathcal{A} \times C)$ with the prohibition of reflexive transitions of the type (i, i) for $i \in C$. In Figure 1, $\mathcal{T} = \{(B, S), (S, E), (E, I), (I, R), (S, D), (E, D), (I, D), (R, D)\}$.

We consider a collection of counting processes $\{\mathbf{N}(t)\} = \{N_{ij}(t) : t \in \mathbb{R}^+, (i, j) \in \mathcal{T}\}$ where $\{N_{ij}(t)\}$ counts events of the ij-type. We now define a process $\{\boldsymbol{X}(t)\} \equiv \{X_c(t) : c \in C, t \in \mathbb{R}^+\}$ by

$$X_c(t) = X_c(0) + \sum_{(i,c)\in\mathcal{T}} N_{ic}(t) - \sum_{(c,j)\in\mathcal{T}} N_{cj}(t). \tag{10}$$

We suppose that $\{\mathbf{X}(t), \mathbf{N}(t)\}$ is a continuous-time Markov chain defined by the transition rates

$$\begin{aligned} q(\boldsymbol{x}, \boldsymbol{k}) &\equiv Q[(x_1, \ldots, x_c), (x_1 + u_1, \ldots, x_c + u_c)] \\ &\equiv \lim_{h \downarrow 0} \frac{P(\boldsymbol{N}(t + h)=\boldsymbol{n} + \boldsymbol{k},\ \boldsymbol{X}(t + h)=\boldsymbol{x} + \boldsymbol{u}|\boldsymbol{X}(t)=\boldsymbol{x}, \boldsymbol{N}(t)=\boldsymbol{n})}{h} \end{aligned} \tag{11}$$

where $\boldsymbol{k} = \{k_{ij}, (i, j) \in \mathcal{T}\} \in \mathbb{N}^{\mathcal{T}}$ and $\boldsymbol{u} = \{u_c, c \in C\}$ with $u_c = \sum_{(i,c)\in\mathcal{T}} k_{ic} - \sum_{(c,j)\in\mathcal{T}} k_{cj}$. Our notation for transition rates uses lower case q when the arguments are the state of origin and the increments of the counting processes, following our univariate MCP notation. We use upper case Q when the arguments are the states from which and to which the transition occurs. Note that the rates in (11) are assumed to depend on $\boldsymbol{x}$ but not $\boldsymbol{n}$, and hence $\{\boldsymbol{X}(t)\}$ is itself a Markov chain. We suppose that all other transition rates for $\{\boldsymbol{X}(t), \boldsymbol{N}(t)\}$ are zero. We call $\{\boldsymbol{X}(t)\}$, constructed in this way, a *Markov counting system* (MCS). Although we have assumed temporal homogeneity when specifying (11), the class of models can readily be extended to include dependence on time. As demonstrated in the measles SEIR example above, it can be convenient to add temporal inhomogeneity after defining stochastic rates in the context of a suitable time homogeneous version of the model.

We call a MCS simple if $q(\boldsymbol{x}, \boldsymbol{k}) > 0$ only when $k_{ij} = 1$ for some $(i, j) \in \mathcal{T}$ and $k_{\ell m} = 0$ for $(\ell, m) \neq (i, j)$. Similarly, we call a MCS compound if $q(\boldsymbol{x}, \boldsymbol{k}) > 0$ only when $k_{ij} > 0$ for some $(i, j) \in \mathcal{T}$ and $k_{\ell m} = 0$ for $(\ell, m) \neq (i, j)$. In words, a simple MCS has only single individuals moving between compartments. A compound MCS is one where simultaneous transitions of the same type are allowed, but simultaneous transitions of different types are not allowed. These two types of MCS may be defined in terms of transition rates $q_{ij}(\boldsymbol{x}, k) \equiv q(\boldsymbol{x}, \mathbf{1}_{ij}k)$ where $\mathbf{1}_{ij}$ is a vector of zeros with a one in the ij position. In general, (11) permits simultaneous events of mixed types, including the bivariate binomial gamma process of Section 3.1.4. Simple MCS models are widely used in science and engineering [17, 25]. Compound MCS processes, interpreted as adding white noise to the rate of a simple MCS, can be constructed by writing $q_{ij}(\boldsymbol{x}, k)$ in terms of the processes developed in Section 3, as illustrated above with the SEIR model. Specifically, if $i \in C$ then one could set $q_{ij}(\boldsymbol{x}, k) = q(x_i, k)$ where $\boldsymbol{x} = \{x_c, c \in C\}$ and $q(x_i, k)$ is the transition rate function for the binomial gamma or binomial beta process. An example of this is the replacement of (7) by (9) to add noise to the transmission rate in the SEIR model. When the components of $\{\boldsymbol{X}(t)\}$ have non-negativity constraints (e.g. the SEIR model, where the number of individuals in each compartment must be non-negative at all times), an unbounded process such as the Poisson gamma process or negative binomial gamma process is inappropriate. When new individuals enter the system, either as immigrants or as newborns, then $i \in \mathcal{A}$. In the case of immigration it might be appropriate to set $q_{ij}(\boldsymbol{x}, k) = q(k)$ where $q(k)$ is the transition rate function of the poisson gamma. Alternatively, modeling new arrivals into the system as a birth process with stochastic rates suggests setting $q_{ij}(\boldsymbol{x}, k) = q(x_j, k)$ where $q(x_j, k)$ is the negative binomial gamma transition rate function. Bivariate dependence, with simultaneous events of mixed type, could be constructed by incorporating a process such as the bivariate binomial gamma process of Section 3.1.4.

The discussion of the SEIR example above might serve as a starting point to formalize a general approach and framework for constructing multivariate systems using simpler lower-dimensional components (including univariate compound MCPs) as 'building blocks' whose transition rates can be 'stacked' to construct MCS models. Some progress in this direction is obtained in Appendix D, and here we summarize these results. We show that, using an appropriate extension of conditions (P1) and (P2), a MCS constructed via blocks inherits the infinitesimal behavior of the individual blocks. We define infinitesimal dispersion of a MCS for each pair $(i, j) \in \mathcal{T}$ in terms of the increments of the counting processes $\{N_{ij}(t)\}$:

$$D_{dX}^{ij}(\boldsymbol{x}) \quad \equiv \quad \frac{\lim_{h\downarrow 0} h^{-1} V[N_{ij}(t+h) - N_{ij}(t) | \boldsymbol{X}(t) = \boldsymbol{x}]}{\lim_{h\downarrow 0} h^{-1} E[N_{ij}(t+h) - N_{ij}(t) | \boldsymbol{X}(t) = \boldsymbol{x}]} \equiv \frac{\sigma_{dX}^{2\,ij}(\boldsymbol{x})}{\mu_{dX}^{ij}(\boldsymbol{x})}$$

Theorem D2 in Appendix D implies that, under general regularity conditions, (i) a simple MCS is necessarily infinitesimally equi-dispersed for all $(i, j) \in \mathcal{T}$; (ii) a compound MCS is infinitesimally over-dispersed for and only for index pairs (i, j) for which $q(\boldsymbol{x}, \boldsymbol{k}) > 0$ for some $\boldsymbol{k}$ with $k_{ij} > 1$; and (iii) for each pair (i, j), the infinitesimal dispersion $D_{dX}^{ij}(\boldsymbol{x})$ is determined only by transition rates involving ij-type transitions. In other words, results for the univariate setting of Sections 2 and 3 extend to multivariate systems of counting processes.

Acknowledgements

This work was supported by National Science Foundation grant DMS-0805533; the RAPIDD program of the Science & Technology Directorate, Department of Homeland Security, the Fogarty International Center, National Institutes of Health; and Spanish Government Projects SEJ2006-03919 and ECO2009-08100 and Spanish Program *Juan de la Cierva* for young researchers. Part of this work was conducted at the Inference for Mechanistic Models Working Group supported by the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #DEB-0553768), the University of California, Santa Barbara, and the State of California. We thank the editor, associate editor and two anonymous referees for their suggestions which led to substantial improvements in the paper.

Appendix A. Proof of Theorem 1

Let $\{P(t)\}$ be a conditional Poisson process with event rate $\bar{\Lambda}(t)$. All probabilities and expectations in the remainder of this proof are conditional on $N(t) = n$ (in addition to other conditioning, where appropriate). Then, since $\Delta N(t)$ is non-negative,

$$E[\Delta N(t)] = E[\Delta N(t)\,\mathbb{I}\{\Delta N(t) > 0\}\,] = E\Big[\,\mathbb{I}\{\Delta N(t) = 1\} + \Delta N(t)\,\mathbb{I}\{\Delta N(t) > 1\}\,\Big]. \tag{A.1}$$

Now, it is immediate that $E[\,\mathbb{I}\{\Delta N(t) = 1\}\,] = \lambda(n)h + o(h)$. Also, since $\{N(t)\}$ is simple, $\{\Delta N(t)\}$ is stochastically smaller than $\{\Delta P(t)\}$ and

$$\begin{aligned} E[\Delta N(t)\,\mathbb{I}\{\Delta N(t) > 1\}\,] &\leq E[\Delta P(t)\,\mathbb{I}\{\Delta P(t) > 1\}\,] \\ &= E\Big[E[\Delta P(t)\,\mathbb{I}\{\Delta P(t) > 1\}\,|\bar{\Lambda}(t)]\Big]. \end{aligned}$$

Using (A.1) with $N(t)$ replaced by $P(t)$, noting also that $E[\Delta P(t)|\bar{\Lambda}(t)] = h\bar{\Lambda}(t)$ and $E[\,\mathbb{I}\{\Delta P(t) = 1\}\,|\bar{\Lambda}(t)] = h\bar{\Lambda}(t)\exp\{-h\bar{\Lambda}(t)\}$, it follows that

$$\begin{aligned} E[\Delta N(t)\,\mathbb{I}\{\Delta N(t) > 1\}\,] &\leq E\Big[h\bar{\Lambda}(t) - h\bar{\Lambda}(t)\exp\{\,-h\bar{\Lambda}(t)\}\Big] \\ &= E\Big[h\bar{\Lambda}(t)\Big(1 - \exp\{-h\bar{\Lambda}(t)\}\Big)\Big]. \end{aligned}$$

It follows by dominated convergence, since $\bar{\lambda}(1 - \exp\{-h\bar{\lambda}\}) \leq \bar{\lambda}$ and by the assumption that $E[\bar{\Lambda}(t)]$ is finite (note that the distribution of $\bar{\Lambda}(t)$ depends on $\bar{h}$ and not h), that

$$\lim_{h\downarrow 0} \frac{E\Big[h\bar{\Lambda}(t)\Big(1 - \exp\{-h\bar{\Lambda}(t)\}\Big)\Big]}{h} \quad = \quad E\Big[\lim_{h\downarrow 0} \bar{\Lambda}(t)\Big(1 - \exp\{-h\bar{\Lambda}(t)\}\Big)\Big] = 0.$$

Therefore, $E[\Delta N(t)\,\mathbb{I}\{\Delta N(t) > 1\}\,] = o(h)$ and $E[\Delta N(t)] = \lambda(n)h + o(h)$. Similarly, replacing first by second moments, $E[(\Delta N(t))^2\,\mathbb{I}\{\Delta N(t) > 1\}\,] = o(h)$ and $E[(\Delta N(t))^2] = \lambda(n)h + o(h)$, since

$$\begin{aligned}
E[(\Delta N(t))^2\,\mathbb{I}\{\Delta N(t) > 1\}\,] &\leq E[(\Delta P(t))^2\,\mathbb{I}\{\Delta P(t) > 1\}\,] \\
&= E\Big[E[(\Delta P(t))^2\,\mathbb{I}\{\Delta P(t) > 1\}\,|\bar{\Lambda}(t)]\Big] \\
&= E[h\bar{\Lambda}(t) + h^2\bar{\Lambda}^2(t) - h\bar{\Lambda}(t)\exp\{-h\bar{\Lambda}(t)\}] \\
&\leq E[2h^2\bar{\Lambda}^2(t)] = o(h),
\end{aligned}$$

where the last line follows by $1 - \exp\{-x\} \leq x$ and $E[\bar{\Lambda}^2(t)]$ being finite. Equi-dispersion follows from $V[\Delta N(t)] = E[(\Delta N(t))^2] - E[\Delta N(t)]^2 = \lambda(n)h + o(h)$, where $E[\Delta N(t)]^2$ is $o(h)$ by stability of $\{N(t)\}$ which implies $\lambda(n) < \infty$ for all n.

Appendix B. Multiplicative Lévy white noise via subordination

Let $\{M_\lambda(t)\}$ be the simple, time homogeneous, conservative and stable MCP with rate function $\lambda : \mathbb{N} \to R^+$ of Theorem 3. It will be convenient here to write $M(\lambda)(t)$ instead of $M_\lambda(t)$. Write $\pi^{M(\lambda)}_{m,m+k}(h)$ for the integrated increment (or transition) probabilities of $\{M(\lambda)(t)\}$, defined as

$$\pi^{M(\lambda)}_{m,m+k}(h) \equiv P\Big(\Delta M(t) = k|M(t) = m\Big).$$

For $\{M(\lambda)(t)\}$, Kolmogorov's Backward Differential System is satisfied [5], i.e.

$$\frac{d}{dh}\pi^{M(\lambda)}_{m,m+k}(h) = \Big[\ \pi^{M(\lambda)}_{m+1,m+k}(h) - \pi^{M(\lambda)}_{m,m+k}(h)\ \Big]\lambda(m). \tag{B.1}$$

This suggests the following definition of $\{M(\lambda\xi)(t)\}$, a simple MCP $\{M(\lambda)(t)\}$ with multiplicative continuous-time noise in the rate function, where $\{\xi(t)\} \equiv \{dL(t)/dt\}$ for a non-decreasing, Lévy integrated noise process $\{L(t)\}$ with $L(0) = 0$ and $E[L(t)] = t$, as in Theorem 3. Define the process $\{M(\lambda\xi)(t)\}$ by

$$\pi^{M(\lambda\xi)}_{m,m+k}(h) \equiv E\Big[\Pi^{M(\lambda\xi)}_{m,m+k}(h)\Big]$$

where $\Pi^{M(\lambda\xi)}_{m,m+k}(h)$ is specified, by analogy to (B.1), as the solution to a stochastic differential equation

$$d\Pi^{M(\lambda\xi)}_{m,m+k}(h) = [\ \Pi^{M(\lambda\xi)}_{m+1,m+k}(h) - \Pi^{M(\lambda\xi)}_{m,m+k}(h)\ ]\lambda(m)\,dL(h), \tag{B.2}$$

or, essentially equivalently,

$$\Pi^{M(\lambda\xi)}_{m,m+k}(h) \quad = \quad \Pi^{M(\lambda\xi)}_{m,m+k}(0) + \int_0^h \Big[\Pi^{M(\lambda\xi)}_{m+1,m+k}(r-) - \Pi^{M(\lambda\xi)}_{m,m+k}(r-)\Big]\lambda(m)\,dL(r) \tag{B.3}$$

To give meaning to (B.2) and (B.3), it is necessary to define a stochastic integral. Here, we use the Marcus canonical stochastic integral with Marcus map $\Phi(u, x, y) = \pi^{M(\lambda)}_{m,m+k}(x + uy)$. The Marcus canonical integral is a stochastic integral developed in the context of Lévy calculus [2]. It is constructed to satisfy a chain rule of the Newton-Leibniz type (unlike the Itô integral). In the case of continuous Lévy processes, the Marcus canonical integral becomes the Stratonovich integral. For jump processes, the Marcus canonical integral heuristically corresponds to approximating trajectories by increasingly accurate continuous piecewise linear functions. We interpret (B.2) as a stochastic version of (B.1). We then think of $\Pi^{M(\lambda\xi)}_{m,m+k}(h)$ as stochastic transition probabilities, conditional on the noise process, giving rise to deterministic transition probabilities $\pi^{M(\lambda\xi)}_{m,m+k}(h)$ once this noise is integrated out.

Proof of Theorem 3*(Lévy white noise and subordination).* By definition

$$\pi^{M(\lambda)\circ L}_{m,m+k}(h) \equiv E\Big[\pi^{M(\lambda)}_{m,m+k}(L(h))\Big].$$

Applying Theorem 4.4.28 of Applebaum [2] with $f(L(h)) = \Pi^{M(\lambda)}_{m,m+k}(L(h))$, it follows that $f \in C^3(\mathbb{R})$ by smoothness of f implied by (B.1). Since $\Pi^{M(\lambda)}_{m,m+k}(0) = 0$, it also follows that

$$\Pi^{M(\lambda)}_{m,m+k}(L(h)) = \int_0^h \Big[\Pi^{M(\lambda)}_{m+1,m+k}(L(r-)) - \Pi^{M(\lambda)}_{m,m+k}(L(r-))\Big]\lambda(m)\, dL(r),$$

so that $\Pi^{M(\lambda)}_{m,m+k}(L(h))$ satisfies (B.3). Given uniqueness and existence of (B.3), we obtain

$$\pi^{M(\lambda)}_{m,m+k}(L(h)) \sim \Pi^{M(\lambda\xi)}_{m,m+k}(h),$$

and hence $\pi^{M(\lambda)\circ L}_{m,m+k}(h) = \pi^{M(\lambda\xi)}_{m,m+k}(h)$. □

Appendix C. Properties of certain compound Markov counting processes

We describe the main steps in the proofs of Propositions 4, 5, 6 and 8. Additional algebraic details are available in Appendix D (Section D2).

Proof of Proposition 4 *(Poisson gamma process).* It is a standard result that if $\alpha\Delta\Gamma(t)$ follows a gamma distribution with mean αh and variance $\alpha^2\tau h$ then the distribution of the increment of $\Delta N(t) = N(t+h) - N(t) = M(\Gamma(t+h)) - M(\Gamma(t))$ is negative binomial with probability mass function

$$P(\Delta N(t) = k | N(t) = n) \quad = \quad \frac{G(\tau^{-1}h + k)}{k!\, G(\tau^{-1}h)} p^{\tau^{-1}h} (1-p)^k \,, \tag{C.1}$$

where $p = (1+\tau\alpha)^{-1}$. The transition rates follow by a Taylor series expansion about $h = 0$, noting that $\frac{G(\eta+k)}{k!\, G(\eta)} = k^{-1}\eta + o(\eta)$. The infinitesimal moments follow directly from the moments of the negative binomial representation in (C.1). □

Proof of Proposition 5 *(binomial gamma process).* Since $\{N(t)\}$ is the counting process associated with a conditional linear death process, the increment process is binomial with parameters size $d_0 - n$ and event probability $\Pi(t) = 1 - e^{-\delta\Delta\Gamma(t)}$, i.e. $P(\Delta N(t) = k | N(t) = n, \Delta\Gamma(t)) = \binom{d_0-n}{k}\Pi(t)^k(1-\Pi(t))^{d_0-n-k}$ for $k \in \{0, 1, \ldots, d_0 - n\}$. We integrate out the continuous-time gamma noise, making use of the multinomial theorem as follows:

$$\begin{aligned}
P(\Delta N = k | N(t) = n) &= \int_0^\infty \binom{d_0-n}{k} [1-e^{-x}]^k [e^{-x}]^{d_0-n-k} \frac{x^{a-1}e^{-xb}b^a}{G(a)} dx \\
&= \binom{d_0-n}{k} \int_0^\infty \left[\sum_{j=0}^k \binom{k}{j} (-e^{-x})^{k-j}\right] e^{-x(d_0-n-k)} \frac{x^{a-1}e^{-xb}b^a}{G(a)} dx \\
&= \binom{d_0-n}{k} \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} \frac{b^a}{(b+d_0-n-j)^a} \times \int_0^\infty \frac{x^{a-1}e^{-x(b+d_0-n-j)}(b+d_0-n-j)^a}{G(a)} dx \\
&= \binom{d_0-n}{k} \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} \Big(1 + \delta\tau(d_0-n-j)\Big)^{-h\tau^{-1}}
\end{aligned}$$

for $k \in \{0, \ldots, d_0 - n\}$, with $a = h\tau^{-1}$ and $b = \delta^{-1}\tau^{-1}$. The transition probabilities follow by applying the Taylor series expansion $(1 + \delta\tau\ell)^{-h\tau^{-1}} = 1 - \tau^{-1}\ln(1 + \delta\tau\ell)h + o(h)$. The integrated moments can be written as

$$
\begin{aligned}
E[\Delta N(t)|N(t) = n] &= (d_0 - n)E[\Pi(t)|N(t) = n] && \text{(C.2)}\\
V[\Delta N(t)|N(t) = n] &= V[(d_0 - n)\Pi(t)|N(t) = n] + E\big[(d_0 - n)\Pi(t)(1 - \Pi(t))|N(t) = n\big] && \text{(C.3)}
\end{aligned}
$$

Recalling that $\Pi(t) = 1 - e^{-\delta\Delta\Gamma(t)}$, and making use of the moment generating function $E\big[\exp\{z\delta\Delta\Gamma(t)\}\big] = (1 - z\delta\tau)^{-\tau^{-1}h}$ for $z\delta\tau < 1$ and $h, \lambda, \tau > 0$, results in a closed form expression for (C.2) and (C.3). The infinitesimal moments follow by taking a Taylor expansion around $h = 0$. □

Proof of Proposition 6 *(negative binomial gamma process).* Since $\{N(t)\}$ is a conditional linear birth process, the increment process is negative binomial with parameters being the number of failures n and the failure probability $\Pi(t) = e^{-\beta\Delta\Gamma(t)}$, i.e.,

$$
P(\Delta N(t) = k|N(t) = n, \Delta\Gamma(t)) \quad = \quad \binom{n + k - 1}{k}\Pi(t)^n\,(1 - \Pi(t))^k\,,
$$

for $k \in \mathbb{N}$. Following a similar calculation to the proof of Proposition 5, with $a = h\tau^{-1}$ and $b = \delta^{-1}\tau^{-1}$, we have

$$
\begin{aligned}
P(\Delta N = k|N(t) = n) = &\int_0^\infty \binom{n + k - 1}{k}\big[e^{-x}\big]^n\big[1 - e^{-x}\big]^k \frac{x^{a-1}e^{-xb}b^a}{G(a)}dx\\
= &\binom{n + k - 1}{k}\int_0^\infty \left[\sum_{j=0}^{k}\binom{k}{j}(-e^{-x})^{k-j}\right]e^{-xn}\frac{x^{a-1}e^{-xb}b^a}{G(a)}dx\\
= &\binom{n + k - 1}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}\Big(1 + \beta\tau(n + k - j)\Big)^{-h\tau^{-1}}
\end{aligned}
$$

The limiting probabilities follow by a Taylor series expansion about $h = 0$. Writing $\Theta(t) = \frac{1-\Pi(t)}{\Pi(t)} = \exp\{\beta\Delta\Gamma(t)\} - 1$, we have

$$
\begin{aligned}
E[\Delta N(t)|N(t) = n] &= nE[\Theta(t)|N(t) = n]\\
V[\Delta N(t)|N(t) = n] &= V[n\Theta(t)|N(t) = n] + E[n\Theta(t)(1 + \Theta(t))|N(t) = n].
\end{aligned}
$$

These moments can be calculated explicitly via the moment generating function, $E[\exp\{z\beta\Delta\Gamma(t)\}] = (1 - z\beta\tau)^{-\tau^{-1}h}$ for $z\beta\tau < 1$ and $h, \lambda, \tau > 0$. The infinitesimal moments follow by taking a Taylor series expansion about $h = 0$. □

Proof of Proposition 8 *(binomial beta process).* $M_h(h) - M_h(0)$ given $M_h(0) = m$ is binomial with parameters size $d_0 - m$ and death probability Π_1, i.e.

$$
P(M_h(h) - M_h(0){=}k|M_h(0){=}m, \Pi_1) = \binom{d_0 - m}{k}(\Pi_1)^k(1 - \Pi_1)^{d_0-m-k}
$$

for $k \in \{0, 1, \ldots, d_0 - m\}$. We integrate out the beta noise using the fact that $M_h(h) - M_h(0)$, conditional on $M_h(0) = m$, has a beta binomial distribution. Setting $a = c(1 - e^{-\delta h})$ and $b = ce^{-\delta h}$, we have

$$
P(M_h(h) - M_h(0){=}k|M_h(0){=}m) = \binom{d_0 - n}{k}\frac{G(a + b)G(k + a)G(d_0 - m - k + b)}{G(a)\,G(b)\,G(a + b + d_0 - m)} \tag{C.4}
$$

for $k \in \{1, \ldots, d_0 - m\}$. Now rewrite (C.4) using the properties that, for $k \geq 1$, $G(k + a) = aG(a)G(k) + o(a)$ and $G(m + b) = \frac{G(m+a+b)G(b)}{G(a+b)} + O(a)$. Noting that $a + b = c$ and that $a = c\delta h + o(h)$, we obtain the required transition probabilities. The infinitesimal moments follow by a Taylor series expansion of the beta binomial moments:

$$
E[M_h(h) - M_h(0)|M_h(0) = m] = (d_0 - m)\frac{a}{a + b}; \qquad V[M_h(h) - M_h(0)|M_h(0) = m] = (d_0 - m)\frac{ab(d_0 - m + a + b)}{(a + b)^2(1 + a + b)}.
$$

The infinitesimal moments of the MCP defined by these limiting transition probabilities can be derived from the moments above because the sums in the conditional moments have a finite number terms $d_0 - m$ and the h limits can therefore be passed inside these moments. $\square$

Appendix D

Submitted to *Stochastic Processes and Their Applications* as an electronic supplement to the preceding article

D1. Properties of certain univariate simple Markov counting processes

Table D1: Increment mean, increment variance and dispersion indices of some standard counting processes. Columns correspond to the time homogeneous Poisson process with intensity α; the linear birth process with per-capita birth rate β and initial population n; the process counting deaths in a linear death process with per-capita death rate δ and initial population d_0.

	Poisson	Birth	Death
$E[\Delta N(t) \mid N(t) = n]$	αh	$n(e^{\beta h} - 1)$	$(d_0 - n)(1 - e^{-\delta h})$
$V[\Delta N(t) \mid N(t) = n]$	αh	$ne^{\beta h}(e^{\beta h} - 1)$	$(d_0 - n)(1 - e^{-\delta h})e^{-\delta h}$
$D_N(n_0, t)$	1	$e^{\beta t}$	$e^{-\delta t}$
$D_{dN}(n)$	1	1	1

The results in Table D1 are all well known. For example, see the discussion of these processes by Bharucha and Reid [4].

D2. Properties of certain univariate compound Markov counting processes

In this section, we give additional details on the proofs of Propositions 4, 5, 6 and 8.

Proof of Proposition 4 *(Poisson gamma process).* Since $\{N(t)\}$ is a conditional Poisson process,

$$P(\Delta N(t) = k \mid N(t), \Delta\Gamma(t)) \quad = \quad \frac{e^{-\alpha\Delta\Gamma(t)}(\alpha\Delta\Gamma(t))^k}{k!}.$$

It is a standard result that if $\alpha\Delta\Gamma(t)$ follows a gamma distribution with mean αh and variance $\alpha^2\tau h$ the distribution of the increments of $\{N(t)\}$ is negative binomial with probability mass function

$$P(\Delta N(t) = k \mid N(t) = n) \quad = \quad \frac{G\left(\tau^{-1}h + k\right)}{k!G(\tau^{-1}h)} p^{\tau^{-1}h} (1-p)^k . \tag{D1}$$

with $p = (1 + \tau\alpha)^{-1}$. The limiting probabilities follow by a Taylor series expansion about $h = 0$:

$$\begin{aligned} P(\Delta N(t) = 0 \mid N(t) = n) \quad &= \quad p^{\tau^{-1}h} = 1 + \tau^{-1}\log(p)h + o(h) \\ P(\Delta N(t) = k \mid N(t) = n) \quad &= \quad \frac{\left(\tau^{-1}h + o(h)\right)}{k} \times \left(1 + \tau^{-1}\log(p)h + o(h)\right)(1-p)^k \\ &= \quad \frac{\tau^{-1}(1-p)^k}{k} h + o(h), \qquad \text{for } k > 0. \end{aligned} \tag{D2}$$

To derive (D2) from (D1) we use the result

$$\begin{aligned} \frac{G(\eta + k)}{k!G(\eta)} \quad &= \quad k!^{-1}(\eta + k - 1) \times (\eta + k - 2) \times \ldots \times (\eta + 2) \times (\eta + 1) \times (\eta) \\ &= \quad \sum_{j=0}^{k-1} k!^{-1}\phi_j h^j \times \eta = k^{-1}\eta + \sum_{j=1}^{k-1} k!^{-1}\phi_j h^{j+1}\tau^{-1} \\ &= \quad k^{-1}\eta + o(h), \end{aligned}$$

with $\eta = \tau^{-1}h$. Recalling that $p = (1+\tau\alpha)^{-1}$, the moments can be derived from standard results:

$$\begin{aligned} E[\Delta N(t)|N(t)=n] &= \frac{\tau^{-1}h(1-p)}{p} \\ &= (1+\tau\alpha)\tau^{-1}h - \tau^{-1}h = \alpha h \\ V[\Delta N(t)|N(t)=n] &= \frac{\tau^{-1}h(1-p)}{p^2} = (1+\tau\alpha)\alpha h. \end{aligned}$$

□

Proof of Proposition 5 *(binomial gamma process).* Since $\{N(t)\}$ is the counting process associated with a conditional linear death process, the increment process is binomial with parameters size $d_0 - n$ and event probability $\Pi(t) = 1 - e^{-\delta\Delta\Gamma(t)}$. Writing $\tilde{n} = d_0 - n$, we have

$$P(\Delta N(t) = k|N(t) = n, \Delta\Gamma(t)) \quad = \quad \binom{\tilde{n}}{k}\Pi(t)^k(1-\Pi(t))^{\tilde{n}-k},$$

for $k \in \{0, 1, \ldots, \tilde{n}\}$. We integrate out the continuous-time gamma noise using the fact that $\delta\Delta\Gamma(t)$ follows a gamma distribution with mean δh and variance $\delta^2\tau h$ (i.e., shape parameters $a = h\tau^{-1}$ and $b = \delta^{-1}\tau^{-1}$). We complete the resulting incomplete gamma density making use of the multinomial theorem as follows

$$\begin{aligned} P(\Delta N = k|N(t) = n) &= \int_0^\infty \binom{\tilde{n}}{k}[1-e^{-x}]^k[e^{-x}]^{\tilde{n}-k}\frac{x^{a-1}e^{-xb}b^a}{G(a)}dx \\ &= \binom{\tilde{n}}{k}\int_0^\infty \left[\sum_{j=0}^k \binom{k}{j}(-e^{-x})^{k-j}\right]e^{-x(\tilde{n}-k)}\frac{x^{a-1}e^{-xb}b^a}{G(a)}dx \\ &= \binom{\tilde{n}}{k}\int_0^\infty \sum_{j=0}^k \binom{k}{j}(-1)^{k-j}e^{-x(\tilde{n}-j)}\frac{x^{a-1}e^{-xb}b^a}{G(a)}dx \\ &= \binom{\tilde{n}}{k}\sum_{j=0}^k \binom{k}{j}(-1)^{k-j}\frac{b^a}{(b+\tilde{n}-j)^a} \times \int_0^\infty \frac{x^{a-1}e^{-x(b+\tilde{n}-j)}(b+\tilde{n}-j)^a}{G(a)}dx \\ &= \binom{\tilde{n}}{k}\sum_{j=0}^k \binom{k}{j}(-1)^{k-j}\Big(1+\delta\tau(\tilde{n}-j)\Big)^{-h\tau^{-1}} \end{aligned}$$

for $k \in \{0, \ldots, \tilde{n}\}$. The limiting probabilities follow by a Taylor series expansion about $h = 0$:

$$\begin{aligned} P(\Delta N(t) = 0|N(t) = n) &= (1+\delta\tau\tilde{n})^{-h\tau^{-1}} = 1 - \tau^{-1}\ln(1+\delta\tau\tilde{n})h + o(h) \\ P(\Delta N(t) = k|N(t) = n) &= \binom{\tilde{n}}{k}\sum_{j=0}^k \binom{k}{j}(-1)^{k-j}\Big(1 - \tau^{-1}\ln(1+\delta\tau(\tilde{n}-j))h + o(h)\Big) \\ &= \binom{\tilde{n}}{k}\sum_{j=0}^k \binom{k}{j}(-1)^{k-j+1}\tau^{-1}\ln(1+\delta\tau(\tilde{n}-j))h + o(h), \end{aligned}$$

for $k \geq 1$, since by the binomial theorem $\sum_{j=0}^k \binom{k}{j}(-1)^{k-j} = (1-1)^k = 0$. The moments are

$$\begin{aligned} E[\Delta N(t)|N(t) = n] &= \tilde{n}E[\Pi(t)|N(t) = n] = \tilde{n}E[1 - e^{-\delta\Delta\Gamma(t)}|N(t) = n] && \text{(D3)} \\ V[\Delta N(t)|N(t) = n] &= V[\tilde{n}\Pi(t)|N(t) = n] + E[\tilde{n}\Pi(t)(1-\Pi(t))|N(t) = n] \\ &= E[\Delta N(t)|N(t) = n] + \tilde{n}\Big[\tilde{n}V[\Pi(t)|N(t) = n] - E[\Pi^2(t)|N(t) = n]\Big] && \text{(D4)} \end{aligned}$$

Let $Y = \delta\Delta\Gamma(t)$. To obtain a closed-form solution for the binomial gamma process, where the probability of death is $\Pi(t) = 1 - e^{-\delta\Delta\Gamma(t)}$, we need $E[e^{-Y}]$, $V[e^{-Y}]$ and $E[(1-e^{-Y})^2]$, which we can get using the moment generating function $E[e^{zY}] = (1 - z\delta\tau)^{-\tau^{-1}h}$ for $z\delta\tau < 1$ and $h, \lambda, \tau > 0$. A Taylor expansion around $h = 0$ then gives

$$\begin{aligned}
E[e^{-Y}] &= (1+\delta\tau)^{-h/\tau} \\
&= 1 - \tau^{-1}\ln(1+\delta\tau)h + o(h) \\
V[e^{-Y}] &= E[e^{-2Y}] - E[e^{-Y}]^2 = (1+2\delta\tau)^{-h/\tau} - (1+\delta\tau)^{-2h/\tau} \\
&= (1 - \tau^{-1}\ln(1+2\delta\tau)h + o(h)) - (1 - \tau^{-1}\ln((1+\delta\tau)^2)h + o(h)) \\
&= \tau^{-1}\ln\Big(\frac{(1+\delta\tau)^2}{1+2\delta\tau}\Big)h + o(h)) \\
E[(1-e^{-Y})^2] &= 1 - 2(1 - \tau^{-1}\ln(1+\delta\tau)h + o(h)) + (1 - \tau^{-1}\ln(1+2\delta\tau)h + o(h)) \\
&= \tau^{-1}\ln\Big(\frac{(1+\delta\tau)^2}{1+2\delta\tau}\Big)h + o(h).
\end{aligned}$$

Plugging these results in the moment expressions in (D3,D4) gives

$$\begin{aligned}
E[\Delta N(t)|N(t)=n] &= \tilde{n}\tau^{-1}\ln(1+\delta\tau)h + o(h) \\
V[\Delta N(t)|N(t)=n] &= \tilde{n}\tau^{-1}\ln(1+\delta\tau)h + \\
&\quad + \tilde{n}\tau^{-1}\Big[(\tilde{n}-1)\ln\Big(\frac{(1+\delta\tau)^2}{1+2\delta\tau}\Big)\Big]h + o(h).
\end{aligned}$$

Since $\frac{(1+\delta\tau)^2}{1+2\delta\tau} > 1$ for $\delta\tau > 0$, it follows that the process is over-dispersed for $\tilde{n} > 1$ and equi-dispersed for $\tilde{n} = 1$. □

Proof of Proposition 7 *(bivariate binomial gamma process).* Letting $\tilde{n}_i = d_{0i} - n_i$ for $i = 1, 2$, it follows by independence of $\{M_1(t)\}$ and $\{M_2(t)\}$ that

$$P(\Delta N_i(t) = k_i, |N_i(t) = n_i, \Delta\Gamma(t)) = \binom{\tilde{n}_1}{k_1}\binom{\tilde{n}_2}{k_2}\Pi(t)^{k_1+k_2}(1-\Pi(t))^{\tilde{n}_1+\tilde{n}_2-k_1-k_2},$$

for $k_i \in \{0, 1, \ldots, \tilde{n}_i\}$. The increment probabilities and transition rates follow respectively from integrating out the continuous-time gamma noise as in the univariate case and from the same Taylor expansions derived in detail above in the proof of proposition 5 in this appendix. The result that the marginal transition rates, increment probabilities and infinitesimal moments are the same as those of the binomial gamma process follows again by independence of $\{M_1(t)\}$ and $\{M_2(t)\}$. Regarding the infinitesimal covariance result, work conditionally on $N_i(t) = n_i$ and define the random variable

$$cov\,[\Delta N_1(t), \Delta N_2(t)|\Pi(t)] \equiv E\,[\Delta N_1(t)\Delta N_2(t)|\Pi(t)] - E\,[\Delta N_1(t)|\Pi(t)]\,E\,[\Delta N_2(t)|\Pi(t)]$$

Then, noting that $cov\,[\Delta N_1(t), \Delta N_2(t)|\Pi(t)]$ is degenerate and equal to zero when $\{M_1(t)\}$ and $\{M_2(t)\}$ are independent, it follows that

$$\begin{aligned}
cov\,[\Delta N_1(t), \Delta N_2(t)] &= cov\,[E\,[\Delta N_1(t)|\Pi(t)]\,, E\,[\Delta N_2(t)|\Pi(t)]] + E\,[cov\,[\Delta N_1(t), \Delta N_2(t)|\Pi(t)]] \\
&= cov\,[\tilde{n}_1\Pi(t), \tilde{n}_2\Pi(t)] \\
&= \tilde{n}_1\tilde{n}_2 V[\Pi(t)] \\
&= \tilde{n}_1\tilde{n}_2\left((1+2\delta\tau)^{-h/\tau} - (1+\delta\tau)^{-2h/\tau}\right) \\
&= \tilde{n}_1\tilde{n}_2\left(\tau^{-1}\ln\Big(\frac{(1+\delta\tau)^2}{1+2\delta\tau}\Big)h + o(h)\right)
\end{aligned}$$

where the expression for $V[\Pi(t)]$ was derived above in the proof of proposition 5 in this appendix. □

Proof of Proposition 6 *(negative binomial gamma process).* Since $\{N(t)\}$ is a conditional linear birth process, the increment process is negative binomial with parameters being the number of failures n and the failure probability $\Pi(t) = e^{-\beta\Delta\Gamma(t)}$, i.e.,

$$P(\Delta N(t) = k|N(t) = n, \Delta\Gamma(t)) = \binom{n+k-1}{k}\Pi(t)^n\,(1-\Pi(t))^k\,,$$

for $k \in \mathbb{N}$. Following a derivation similar to that of the binomial gamma processes of Proposition 5, with $a = h\tau^{-1}$ and $b = \delta^{-1}\tau^{-1}$, we have

$$P(\Delta N = k|N(t) = n) = \int_0^\infty \binom{n+k-1}{k}[e^{-x}]^n[1-e^{-x}]^k \frac{x^{a-1}e^{-xb}b^a}{G(a)}dx$$

$$= \binom{n+k-1}{k}\int_0^\infty \Big[\sum_{j=0}^{k}\binom{k}{j}(-e^{-x})^{k-j}\Big]e^{-xn}\frac{x^{a-1}e^{-xb}b^a}{G(a)}dx$$

$$= \binom{n+k-1}{k}\sum_{j=0}^{k}\binom{k}{j}(-1)^{k-j}\Big(1+b\tau(n+k-j)\Big)^{-h\tau^{-1}}$$

The limiting probabilities follow by a Taylor series expansion about $h = 0$ like in the proof of Proposition 5. The moments can be found as follows. Consider the odds of a birth $\Theta(t) = \frac{1-\Pi(t)}{\Pi(t)}$ given the probability of a birth $1 - \Pi(t)$. Then

$$\begin{aligned}
E[\Delta N(t)|N(t) = n] &= nE[\Theta(t)|N(t) = n] = nE[e^{\beta\Delta\Gamma(t)} - 1|N(t) = n]\\
V[\Delta N(t)|N(t) = n] &= V[n\Theta(t)|N(t) = n] + E[n\Theta(t)(1+\Theta(t))|N(t) = n]\\
&= E[\Delta N(t)|N(t) = n] + n\Big[nV[\Theta(t)|N(t) = n] + E[\Theta^2(t)|N(t) = n]\Big]
\end{aligned}$$

Let $Y = \rho\Delta\Gamma(t)$. Then, Y follows a gamma distribution with mean βh and variance $\beta^2\tau h$. To obtain a closed-form solution for the binomial gamma process, where the odds of a birth is $\Theta(t) = e^{\beta\Delta\Gamma(t)} - 1$, we need $E[e^Y]$, $V[e^Y]$ and $E[(e^Y-1)^2]$, which we can get using the moment generating function $E[e^{zY}] = (1 - z\delta\tau)^{-\tau^{-1}h}$ for $z\delta\tau < 1$ and $h, \lambda, \tau > 0$. Via a Taylor expansion around $h = 0$, we obtain

$$\begin{aligned}
E[e^Y] &= (1-\beta\tau)^{-h/\tau}\\
&= 1 - \tau^{-1}\ln(1-\beta\tau)h + o(h)\\
V[e^Y] &= E[e^{2Y}] - E[e^Y]^2 = (1-2\beta\tau)^{-h/\tau} - (1-\beta\tau)^{-2h/\tau}\\
&= \{1 - \tau^{-1}\ln(1-2\beta\tau)h + o(h)\} - \{1 - \tau^{-1}\ln((1-\beta\tau)^2)h + o(h)\}\\
&= \tau^{-1}\ln\Big(\frac{(1-\beta\tau)^2}{1-2\beta\tau}\Big)h + o(h)\\
E[(e^Y-1)^2] &= 1 - 2\{1 - \tau^{-1}\ln(1-\beta\tau)h + o(h)\} + \{1 - \tau^{-1}\ln(1-2\beta\tau)h + o(h)\}\\
&= \tau^{-1}\ln\Big(\frac{(1-\beta\tau)^2}{1-2\beta\tau}\Big)h + o(h).
\end{aligned}$$

Note that we require that $2\tau\beta < 1$. Plugging this into the moment expressions gives

$$\begin{aligned}
E[\Delta N(t)|N(t)n = n] &= n\tau^{-1}\ln\Big(\frac{1}{1-\beta\tau}\Big)h + o(h)\\
V[\Delta N(t)|N(t) = n] &= n\tau^{-1}\ln\Big(\frac{1}{1-\beta\tau}\Big)h +\\
&\quad + n\tau^{-1}\Big[(n-1)\ln\Big(\frac{(1-\beta\tau)^2}{1-2\beta\tau}\Big)\Big]h + o(h).
\end{aligned}$$

Since $\frac{(1-\beta\tau)^2}{1-2\beta\tau} > 1$ for $\beta\tau > 0$ and $2\tau\beta < 1$, it follows that the process is also over-dispersed for $n > 1$ and equi-dispersed for $n = 1$. □

Proof of Proposition 8 *(binomial beta process).* $M_h(h) - M_h(0)$ given $M_h(0) = m$ is binomial with parameters size $d_0 - m$ and death probability Π_1, i.e., setting $\tilde{m} = d_0 - m$, we have

$$P(M_h(h) - M_h(0){=}k|M_h(0){=}m, \Pi_1) = \binom{\tilde{m}}{k}(\Pi_1)^k(1-\Pi_1)^{\tilde{m}-k},$$

for $k \in \{0, 1, \dots, \tilde{m}\}$. We integrate out the beta noise using the fact that $M_h(h) - M_h(0)$ conditional on $M_h(0) = d_0 - m$ has a beta binomial distribution with the corresponding parameters. The beta binomial probability mass function of $M_h(h) - M_h(0)$ given $M_h(0) = m$ is

$$\begin{aligned}
& P(M_h(h) - M_h(0)=k|M_h(0)=m) = \\
= & \binom{\tilde{m}}{k} \frac{G(a+b)G(k+a)G(\tilde{m}-k+b)}{G(a)G(b)G(a+b+\tilde{m})} && \text{(D5)} \\
= & \binom{\tilde{m}}{k} \frac{G(a+b)G(a)G(b)G(k)a\frac{G(c+\tilde{m}-k)}{G(c)}}{G(a+b)G(a)G(b)\frac{G(c+\tilde{m})}{G(c)}} + o(h) && \text{(D6)} \\
= & \binom{\tilde{m}}{k} \frac{G(k)G(c+\tilde{m}-k)}{G(c+\tilde{m})} c\delta h + o(h), && \text{(D7)}
\end{aligned}$$

for $k \in \{1, \dots, \tilde{m}\}$. For $k = \tilde{m}$, (D6) follows by $G(k+a) = aG(a)G(k) + o(a)$, which holds for $k \geq 1$. For $k < \tilde{m}$, an application of Lemma D1 in this appendix is also needed. Specifically, using Lemma D1 with $i = \tilde{m} - k$, it follows that

$$G(\tilde{m} - k + b) = \Big\{\frac{G(c+\tilde{m}-k)}{G(c)} + O(h)\Big\} G(b), \tag{D8}$$

and, since $a + b = c$,

$$\begin{aligned}
G(a+b+\tilde{m}) &= G(c+\tilde{m}) && \text{(D9)} \\
&= \frac{G(c+\tilde{m})}{G(c)} G(c) \\
&= \frac{G(c+\tilde{m})}{G(c)} G(a+b).
\end{aligned}$$

Plugging (D8) and (D9) into (D5) gives (D6). Then, using $a = c\delta h + o(h)$ and canceling terms gives (D7), which corresponds to the transition rates. The moments of a beta binomial distribution are a standard result. Since $a = c(1 - e^{-\delta h})$ and $b = ce^{-\delta h}$ and $c = \omega^{-1}(\tilde{m} - 1) - 1$ for $\tilde{m} > 1$, Taylor expansions around $h = 0$ then give

$$\begin{aligned}
E[M_h(h) - M_h(0)=k|M_h(0)=m] &= \tilde{m}\frac{a}{a+b} \\
&= \tilde{m}\delta h + o(h) \\
V[M_h(h) - M_h(0)=k|M_h(0)=m] &= \tilde{n}\frac{ab}{(a+b)^2}\frac{\tilde{n}+a+b}{1+a+b} \\
&= \tilde{m}(1-e^{-\delta h})e^{-\delta h}\frac{\tilde{n}+c}{c+1} \\
&= \tilde{m}\delta h\,(1+\omega) + o(h),
\end{aligned}$$

for $\tilde{m} > 1$ and it follows that the binomial beta process is over-dispersed for $\omega > 0$. If $\tilde{m} = 1$ the process is equi-dispersed as

$$\begin{aligned}
V[M_h(h) - M_h(0)=k|M_h(0)=m] &= \tilde{m}\frac{ab}{(a+b)^2} \\
&= \tilde{m}\delta h + o(h).
\end{aligned}$$

The infinitesimal moments of the MCP defined by these limiting transition probabilities can be derived from the moments above because the sums in the conditional moments have a finite number terms $d_0 - m$ and the h limits can therefore be passed inside these moments. □

Lemma D1. *For $a = c(1 - e^{-\delta h})$, $b = ce^{-\delta h}$, $c > 0$ and $i \in \{1, 2, \dots\}$,*

$$G(b+i) = \Big\{\frac{G(c+i)}{G(c)} + O(h)\Big\} G(b).$$

Proof. Since $b = c - a$, and by the definition of the gamma function, for $i \geq 1$,

$$\begin{aligned} G(b+i) &= (c-a+(i-1)) \times (c-a+(i-2)) \times \cdots \times (c-a) \times G(b) \\ &= \{(c+(i-1)) \times (c+(i-2)) \times \cdots \times (c) + O(h)\} G(b) \\ &= \Big\{ \prod_{j=0}^{i-1} (c+j) + O(h) \Big\} G(b) \\ &= \Big\{ \frac{G(c+i)}{G(c)} + O(h) \Big\} G(b). \end{aligned}$$

□

D3. Equi-dispersion of mixed simple Markov counting processes

In this section, we present a proof of Theorem 2.

Proof of Theorem 2 *(sufficient condition for mixed Markov infinitesimal equi-dispersion).* We proceed similarly to the proof of Theorem 1. Now that $\{N(t)\}$ is a mixed MCP, the bound $\bar{\Lambda}^*(t)$ which features in P1* and P2* is stochastic both because of its dependence on $\{N(t)\}$ and its dependence on the random variable M. All probabilities and expectations in the remainder of this proof are conditional on $N(t) = n$ (in addition to other conditioning, where appropriate). The result for the mean follows as in Theorem 1 by dominated convergence but the dominated functions are now

$$E\Big[\bar{\Lambda}^*(t)\Big(1 - \exp\{-h\bar{\Lambda}^*(t)\}\Big)\Big|M = m\Big] \leq E\Big[\bar{\Lambda}^*(t)\Big|M = m\Big],$$

for all m, and the dominating function has a finite integral since $E\Big[E[\bar{\Lambda}^*(t)|M]\Big] = E[\bar{\Lambda}^*(t)] < \infty$. Then,

$$\lim_{h\downarrow 0} \frac{E[E[h\bar{\Lambda}^*(t)(1 - \exp\{-h\bar{\Lambda}^*(t)\})|M]]}{h} = E[\lim_{h\downarrow 0} \bar{\Lambda}^*(t)(1 - \exp\{-h\bar{\Lambda}^*(t)\})] = 0.$$

The result for the variance follows again in the same lines as for the non-mixing case, i.e.

$$\begin{aligned} E[(\Delta N(t))^2\, \mathbb{I}\{\Delta N(t) > 1\}\,] &= E[E[(\Delta N(t))^2\, \mathbb{I}\{\Delta N(t) > 1\}\,|M]] \\ &\leq 2h^2 E[E[\bar{\Lambda}^{*\,2}(t)|M]] = o(h), \end{aligned}$$

by the assumption that $E[\bar{\Lambda}^{*\,2}(t)] < \infty$. These two results show that the same terms that vanished in Theorem 1 vanish now as well. Then,

$$\begin{aligned} E[\Delta N(t)] = E[\,\mathbb{I}\{\Delta N(t) = 1\}\,] + o(h) &= E[E[\,\mathbb{I}\{\Delta N(t) = 1\}\,|M]] + o(h) \\ &= E[\Lambda(n)\phi^*(h)] + o(h) \qquad \text{(D10)} \end{aligned}$$

where (D10) follows by Lemma D3. Here

$$\phi^*(h) = \begin{cases} \exp\{-h\Lambda(n+1)\}h & \text{if } \Lambda(n) = \Lambda(n+1) \\ \dfrac{\exp\{-h\Lambda(n+1)\} - \exp\{-h(\Lambda(n)\}}{\Lambda(n) - \Lambda(n+1)} & \text{if } \Lambda(n) \neq \Lambda(n+1) \end{cases}$$

Taking limits, as in (D21), gives the desired result via dominated convergence:

$$\lim_{h\downarrow 0} \frac{E[\Delta N(t)]}{h} = \lim_{h\downarrow 0} E\left[\Lambda(n)\frac{\phi^*(h)}{h}\right] = E\left[\Lambda(n) + \lim_{h\downarrow 0} \frac{o_M(h)}{h}\right] = E[\Lambda(n)],$$

where it follows analogously to (D12) that $\phi^*(h)/h \leq 1$ and thus $\Lambda(n)$ dominates $\Lambda(n)\phi^*(h)/h$ with $E[\Lambda(n)] \leq E[\bar{\Lambda}^*(t)] < \infty$. Here, as in Lemma D3, $o_M(h)$ terms are standard $o(h)$ terms for every fixed valued m of the random variable M. The same argument gives $\lim_{h\downarrow 0} h^{-1}E[(\Delta N(t))^2] = E[\Lambda(n)]$ and the same dispersion results as in Theorem 1 follow. □

D4. Sufficient and necessary conditions for equi- and over-dispersion of Markov counting systems

In this section, we generalize the results of the univariate Section 2 to the multivariate Markov counting system framework of Section 4. Analogously to Section 2, define the rate function of a MCS to be

$$\lambda(\boldsymbol{x}) \equiv \lim_{h\downarrow 0} \frac{1 - P(\Delta N_{ij} = 0 \text{ for all } (i, j)|\boldsymbol{X}(t) = \boldsymbol{x})}{h}.$$

Again we restrict ourselves to stable and conservative processes so that $\lambda(\boldsymbol{x}) = \sum_{\boldsymbol{k}} q(\boldsymbol{x}, \boldsymbol{k}) < \infty$ for all $\boldsymbol{x}$. We introduce two regularity conditions, analogous to those of Section 2. Here, the stochastic bound on the rate function is now

$$\bar{\Lambda}_{ij}(t) \equiv \sup_{t\le s\le t+\bar{h}} \lambda(\boldsymbol{X}(s)).$$

Since, unlike in the result in the univariate section, we are now allowing for compound processes, we also need a stochastic bound on the size of increments of the counting processes,

$$\bar{Z}_{ij}(t) \equiv \sup_{t\le s\le t+\bar{h}} dN_{ij}(s).$$

Then, the conditions are

P3. For each t, $\boldsymbol{x}$ and $i \neq j$ there is some $\bar{h} > 0$ such that $E[\bar{Z}_{ij}(t)\bar{\Lambda}_{ij}(t)|\boldsymbol{X}(t) = \boldsymbol{x}] < \infty$.

P4. For each t, $\boldsymbol{x}$ and $i \neq j$ there is some $\bar{h} > 0$ such that $V[\bar{Z}_{ij}(t)\bar{\Lambda}_{ij}(t)|\boldsymbol{X}(t) = \boldsymbol{x}] < \infty$.

Properties P3 and P4 are multivariate extensions of P1 and P2 respectively, requiring that the ij-marginal counting processes do not have an explosive behavior. In particular, P3 and P4 hold for simple birth-death processes with linear birth and death rates. When P3 and P4 hold, Theorem D2 shows it is sufficient and necessary that all ij-marginal processes $\{N_{ij}(t)\}$ associated with $\{\boldsymbol{X}(t)\}$ be simple for the latter to be infinitesimally equi-dispersed. Sufficiency follows because for simple ij-marginal processes $\{N_{ij}(t)\}$,

$$D^{ij}_{d\boldsymbol{X}}(\boldsymbol{x}) \quad = \quad \frac{\sum_k k^2 q_{ij}(\boldsymbol{x}, k)}{\sum_k k q_{ij}(\boldsymbol{x}, k)} = \frac{q_{ij}(\boldsymbol{x}, 1)}{q_{ij}(\boldsymbol{x}, 1)} = 1.$$

Necessity follows because for compound ij-marginal processes $\{N_{ij}(t)\}$,

$$D^{ij}_{d\boldsymbol{X}}(\boldsymbol{x}) \quad = \quad \frac{\sum_k k^2 q_{ij}(\boldsymbol{x}, k)}{\sum_k k q_{ij}(\boldsymbol{x}, k)} > 1$$

since $k^2 > k$ for $k > 1$. Sufficiency and necessity of compoundness for over-dispersion follow analogously. In this appendix, we define $q_{ij}(\boldsymbol{x}, \ell) \equiv \sum_{\boldsymbol{k}:k_{ij}=\ell} q(\boldsymbol{x}, \boldsymbol{k})$ to be the *marginal transition rate* function for the ij-type transition of a MCS. This notation generalizes and supersedes the notation in the main text, where $q_{ij}(\boldsymbol{x}, \ell)$ was defined to be $q(\boldsymbol{x}, \mathbf{1}_{ij}\ell)$. The previous, simpler notation was adequate for the main text since there we focused on compound and simple processes for which simultaneous transitions of different types were not allowed.

Theorem D2 (infinitesimal moments of a MCS)**.** *Let $\{\boldsymbol{X}(t)\}$ be a time homogeneous, stable and conservative Markov counting system with associated multivariate counting process $\{\boldsymbol{N}(t)\}$ as defined in the main text by (10) and (11). Supposing (P3), the infinitesimal mean of $\{N_{ij}(t)\}$ is $\mu^{ij}_{d\boldsymbol{X}}(\boldsymbol{x}) = \sum_k k q_{ij}(\boldsymbol{x}, k)$. Supposing (P4), its infinitesimal variance is $\sigma^{2\,ij}_{d\boldsymbol{X}}(\boldsymbol{x}) = \sum_k k^2 q_{ij}(\boldsymbol{x}, k)$.*

Proof. Let $\{\bar{N}_{ij}(t)\}$ be a conditional compound Poisson process with event rate $\bar{\Lambda}(t) \equiv \bar{\Lambda}_{ij}(t)$ and degenerate jump distribution with mass one at $\bar{Z}(t) \equiv \bar{Z}_{ij}(t)$. All probabilities and expectations in the remainder of this proof are conditional on $N(t) = n$ (in addition to other conditioning, where appropriate). Let S be the event that there is exactly one transition time occurring in the interval $[t, t+h]$, as in Lemma D3. This transition may involve increments in one or more of the $\{N_{ij}(t)\}$ processes, and these increments may be of size one or more. Then,

$$E[\Delta N_{ij}(t)] \;=\; E[\Delta N_{ij}(t)\,\mathbb{I}\{S\}] + E[\Delta N_{ij}(t)\,\mathbb{I}\{S^c\}]. \tag{D11}$$

Unlike in Theorem 1, the term corresponding to one single transition time is not immediate and requires approximating $P(S|\boldsymbol{X}(t)=\boldsymbol{x})$. Lemma D3 provides us with such a result, namely $P(S|\boldsymbol{X}(t)=\boldsymbol{x}) = h\lambda(\boldsymbol{x}) + o(h)$. Letting S_{ij} be the event that there is exactly one transition time occurring in the interval $[t, t+h]$ and that this transition involves increments in the $\{N_{ij}(t)\}$ process (and also possibly in other processes), we write

$$\begin{aligned} E[\Delta N_{ij}(t)\,\mathbb{I}\{S\}] &= E[\Delta N_{ij}(t)|S_{ij}, \boldsymbol{X}(t)=\boldsymbol{x}] \times P(S_{ij}|S, \boldsymbol{X}(t)=\boldsymbol{x}) \times P(S|\boldsymbol{X}(t)=\boldsymbol{x}) \\ &= \sum_k k\frac{q_{ij}(\boldsymbol{x},k)}{\sum_k q_{ij}(\boldsymbol{x},k)} \times \frac{\sum_k q_{ij}(\boldsymbol{x},k)}{\lambda(\boldsymbol{x})} \times \Big[h\lambda(\boldsymbol{x}) + o(h)\Big] \\ &= h\sum_k k q_{ij}(\boldsymbol{x},k) + o(h). \end{aligned}$$

Analogously to Theorem 1, we proceed to bound the second term to show the desired result. Since $\Delta N_{ij}(t)$ is stochastically smaller than $\Delta\bar{N}_{ij}(t)$,

$$\begin{aligned} E[\Delta N_{ij}(t)\,\mathbb{I}\{S^c\}] &\leq E[\Delta\bar{N}_{ij}(t)\,\mathbb{I}\{S^c\}] \\ &= E\Big[E[\Delta\bar{N}_{ij}(t)\,\mathbb{I}\{S^c\}\,|\bar{\Lambda}(t), \bar{Z}(t)]\Big] \end{aligned}$$

Using (D11) with $N_{ij}(t)$ replaced by $\bar{N}_{ij}(t)$ and since $E[\Delta\bar{N}_{ij}(t)|\bar{\Lambda}(t), \bar{Z}(t)] = \bar{Z}(t)h\bar{\Lambda}(t)$ and $E[\Delta\bar{N}_{ij}(t)\,\mathbb{I}\{S\}\,|\bar{\Lambda}(t), \bar{Z}(t)] = \bar{Z}(t)h\bar{\Lambda}(t)\exp\{-h\bar{\Lambda}(t)\}$, it follows that

$$\begin{aligned} E[\Delta N_{ij}(t)\,\mathbb{I}\{S^c\}] &\leq E\Big[\bar{Z}(t)h\bar{\Lambda}(t) - \bar{Z}(t)h\bar{\Lambda}(t)\exp\{-h\bar{\Lambda}(t)\}\Big] \\ &= E\Big[\bar{Z}(t)h\bar{\Lambda}(t)(1-\exp\{-h\bar{\Lambda}(t)\})\Big]. \end{aligned}$$

As in Theorem 1, it follows by dominated convergence, since $\bar{z}\bar{\lambda}\Big(1-\exp\{-h\bar{\lambda}\}\Big) \leq \bar{z}\bar{\lambda}$ and $E[\bar{Z}(t)\bar{\Lambda}(t)]$ is finite, that

$$\lim_{h\downarrow 0}\frac{E\Big[\bar{Z}(t)h\bar{\Lambda}(t)\Big(1-\exp\{-h\bar{\Lambda}(t)\}\Big)\Big]}{h} \;=\; E\Big[\lim_{h\downarrow 0}\bar{Z}(t)\bar{\Lambda}(t)\Big(1-\exp\{-h\bar{\Lambda}(t)\}\Big)\Big] = 0.$$

Therefore, $E[\Delta N_{ij}(t)\,\mathbb{I}\{S^c\}] = o(h)$ and the result for the mean follows. Replacing first by second moments, the result for the variance follows since

$$\begin{aligned} E[(\Delta N_{ij}(t))^2\,\mathbb{I}\{S^c\}] &\leq E[(\Delta\bar{N}_{ij}(t))^2\,\mathbb{I}\{S^c\}] \\ &= E\Big[E[(\Delta\bar{N}_{ij}(t))^2\,\mathbb{I}\{S^c\}\,|\bar{\Lambda}(t), \bar{Z}(t)]\Big] \\ &= E[\bar{Z}^2(t)h\bar{\Lambda}(t) + \bar{Z}^2(t)h^2\bar{\Lambda}^2(t) - \bar{Z}^2(t)h\bar{\Lambda}(t)\exp -h\bar{\Lambda}(t)] \\ &\leq E[2\bar{Z}^2(t)h^2\bar{\Lambda}^2(t)] = o(h), \end{aligned}$$

since $E[\bar{Z}^2(t)\bar{\Lambda}^2(t)]$ is assumed to be finite. □

To complete the proof of Theorem D2, we require the following lemma. This technical result is similar to, but slightly different from, standard results on Markov chains. Equation (D13) was shown, for example, by [32, page 492]. However, the inequality in (D12) is, to our knowledge, new.

Lemma D3 (probability of a single event time in Markov counting systems)**.** *Let $\{\boldsymbol{X}(t)\}$ be a time homogeneous, stable and conservative Markov counting system with associated multivariate counting process $\{\boldsymbol{N}(t)\}$ as defined in the main text by (11) and (10). Consider a starting time t and let U be the time between t and the first event time and*

V be the time between $t + U$ and the second event time. Let S be the event that there is exactly one transition time occurring in the interval $[t, t+h]$. This transition may involve increments in one or more of the $\{N_{ij}(t)\}$ processes, and these increments may be of size one or more. Then, letting $\lambda_U \equiv \lambda(\boldsymbol{x})$ be the rate function of $\{\boldsymbol{X}(t)\}$ during $[t, t+U]$ and $\Lambda_V \equiv \lambda(\boldsymbol{X}(t+U))$ be this conditional rate function during $[t+U, t+U+V]$,

$$P(S\,|\boldsymbol{X}(t) = \boldsymbol{x}, \Lambda_V) = \lambda_U \phi(h) \quad \leq \quad \lambda_U h \tag{D12}$$

where

$$\phi(h) \equiv \begin{cases} e^{-h\Lambda_V} h & \text{if } \lambda_U = \Lambda_V \\ \frac{e^{-h\Lambda_V} - e^{-h\lambda_U}}{\lambda_U - \Lambda_V} & \text{if } \lambda_U \neq \Lambda_V \end{cases}$$

and

$$P(S\,|\boldsymbol{X}(t) = \boldsymbol{x}) = h\lambda(\boldsymbol{x}) + o(h). \tag{D13}$$

Proof. Start by fixing the random variable Λ_V at a given constant, say λ_V. Given the Markov property, for the starting time t, the densities of the exponential inter-event times are $f_U(u) = \lambda_U \mathrm{e}^{-u\lambda_U}$ for $u > 0$ and $f_V(v) = \lambda_V \mathrm{e}^{-v\lambda_V}$ for $v > 0$. Then,

$$\begin{aligned}
P(S\,|\boldsymbol{X}(t) = \boldsymbol{x}) &= P(U < h, U + V > h) = P(U < h, V > h - U) \\
&= \int_0^h \int_{h-u}^{\infty} f_{U,V}(u, v) dv du = \int_0^h \int_{h-u}^{\infty} \lambda_U \mathrm{e}^{-u\lambda_U} \lambda_V \mathrm{e}^{-v\lambda_V} dv du \\
&= \int_0^h \lambda_U \mathrm{e}^{-u\lambda_U} \lambda_V du \int_{h-u}^{\infty} \mathrm{e}^{-v\lambda_V} dv = \int_0^h \lambda_U \mathrm{e}^{-u\lambda_U} \mathrm{e}^{-(h-u)\lambda_V} du \\
&= \lambda_U \mathrm{e}^{-h\lambda_V} \int_0^h \mathrm{e}^{-u(\lambda_U - \lambda_V)} du.
\end{aligned} \tag{D14}$$

If the event rate is not changed by the first event happening (like it happens in a Poisson process but unlike in linear birth or death processes), then we can write $\lambda_U = \lambda_V = \lambda$ in (D14) and

$$\begin{aligned}
P(S\,|\boldsymbol{X}(t) = \boldsymbol{x}, \lambda_V) &= \lambda_U \mathrm{e}^{-h\lambda_V} h && \text{(D15)} \\
&= \lambda h (1 - \lambda h + o(h)) \\
&= \lambda h + o(h). && \text{(D16)}
\end{aligned}$$

If $\lambda_U \neq \lambda_V$, then from (D14)

$$\begin{aligned}
P(S\,|\boldsymbol{X}(t) = \boldsymbol{x}, \lambda_V) &= \lambda_U \mathrm{e}^{-h\lambda_V} \left[\frac{1 - \mathrm{e}^{-h(\lambda_U - \lambda_V)}}{\lambda_U - \lambda_V} \right] \\
&= \lambda_U \frac{\mathrm{e}^{-h\lambda_V} - \mathrm{e}^{-h\lambda_U}}{\lambda_U - \lambda_V} && \text{(D17)} \\
&= \lambda_U \frac{1 - h\lambda_V + o(h) - 1 + h\lambda_U + o(h)}{\lambda_U - \lambda_V} \\
&= \lambda_U h \frac{\lambda_U - \lambda_V}{\lambda_U - \lambda_V} + o(h) \\
&= \lambda_U h + o(h). && \text{(D18)}
\end{aligned}$$

Combining (D15) and (D17), replacing λ_V by Λ_V and conditioning on Λ_V gives the equality in (D12). The inequality in (D12) follows directly for the case $\lambda_U = \Lambda_V$ since $\exp\{-h\Lambda_V\} \leq 1$. For $\lambda_U \neq \Lambda_V$, consider $f(x) = \exp\{-x\}$ with $x \in \mathbb{R}^+$. The mean value theorem asserts that, for some non-negative real $z \in [x, y]$,

$$f(y) = f(x) + (y - x) df/dx(z). \tag{D19}$$

Rearranging (D19), we obtain

$$\frac{\exp\{-y\} - \exp\{-x\}}{(x-y)} = \exp\{-z\}. \tag{D20}$$

The inequality in (D12) follows by setting $x = h\lambda_U$ and $y = h\Lambda_V$ in (D20), noting that $\exp\{-z\} \in [0,1]$. We now complete the proof by showing that equation (D13) follows from (D16) and (D18) via dominated convergence, since

$$\lim_{h\downarrow 0} h^{-1} P(S\,|\boldsymbol{X}(t)=\boldsymbol{x}) = \lim_{h\downarrow 0} h^{-1} E[P(S\,|\boldsymbol{X}(t)=\boldsymbol{x}, \Lambda_V)] \quad = \quad E[\lim_{h\downarrow 0} h^{-1}(\lambda_U h + o_{\Lambda_V}(h))] = \lambda_U. \tag{D21}$$

Here, $o_{\Lambda_V}(h)$ terms are standard $o(h)$ terms for every fixed valued λ_V of the random variable Λ_V. To justify passing the h limit inside the expected value, note that $P(S\,|\boldsymbol{X}(t)=\boldsymbol{x})/h$ is dominated by λ_U from (D12). □

D5. Construction of over-dispersed Markov counting systems

Extending the results of Section 3 to the multivariate Markov counting systems of Section 4 requires considering infinitesimally over-dispersed MCS models. A direct extension of Section 3 could lead to subordination of multivariate infinitesimally equi-dispersed processes. However, in Section 4, we have already introduced the idea of stacking blocks of small dimension to create larger systems. The key question is whether processes built following this approach inherit the infinitesimal properties of the smaller blocks. Theorem D2 from Section D4 gives sufficient conditions for this, which in turn implies that the system will be over-dispersed if built with over-dispersed univariate blocks. In this section we give an additional example of the 'building block' strategy by constructing in Proposition D4 an infinitesimally over-dispersed birth-death process by stacking the transition rates of the binomial gamma and of the negative binomial gamma process of Section 3. Then, we check that infinitesimal dispersion of the building processes is retained by this birth-death process and by the models from the SEIR example of Section 4 by checking that P3 and P4 hold.

Proposition D4 (infinitesimally over-dispersed birth-death process)**.** *Consider one compartment $C = \{Y\}$ and two auxiliary nodes $\mathcal{A} = \{B, D\}$ with allowed transitions $\mathcal{T} = \{(B,Y),(Y,D)\}$ representing births ($B \to Y$) and deaths ($Y \to D$). Based on these, define the compound Markov counting system $\{Y(t)\}$ by*

$$Y(t) = Y(0) + N_{BY}(t) - N_{YD}(t) \tag{D22}$$

and by birth transition rates, for $\boldsymbol{k} = \{k_{BY}, 0\}$ with $k_{BY} \geq 1$,

$$q(y,\boldsymbol{k}) \equiv q_{BY}(y,k_{BY}) \quad = \quad \binom{y+k_{BY}-1}{k_{BY}} \sum_{j=0}^{k_{BY}} \binom{k_{BY}}{j} (-1)^{k-j+1} (\tau^B)^{-1} \ln\left(1 + \beta\tau^B (y + k_{BY} - j)\right)$$

and by death transition rates, for $\boldsymbol{k} = \{0, k_{YD}\}$ with $k_{YD} \in \{0,\dots,y\}$,

$$q(y,\boldsymbol{k}) \equiv q_{YD}(y,k_{YD}) \quad = \quad \binom{y}{k_{YD}} \sum_{j=0}^{k_{YD}} \binom{k_{YD}}{j} (-1)^{k_{YD}-j+1} (\tau^D)^{-1} \ln\left(1 + \delta\tau^D (y-j)\right)$$

for $y > 0$ and zero otherwise. Then, the infinitesimal dispersion of such Markov counting system is given by

$$\begin{aligned} D^{BY}_{dY}(y) \quad &= \quad 1 + (y-1)\left[\frac{2\ln(1-\beta\tau^B) - \ln(1-2\beta\tau^B)}{-\ln(1-\beta\tau^B)}\right] \\ D^{YD}_{dY}(y) \quad &= \quad 1 + (y-1)\left[\frac{2\ln(1+\delta\tau^D) - \ln(1+2\delta\tau^D)}{\ln(1+\delta\tau^D)}\right]. \end{aligned}$$

for $2\beta\tau^B < 1$ and provided $Y(0) > 0$. Hence, the ij-marginals of the counting process $\{\boldsymbol{N}(t)\}$ associated to the MCS $\{Y(t)\}$ are infinitesimally over-dispersed for $y > 1$ and equi-dispersed for $y = 1$.

Proof. The results follow by Theorem D2. Hence, we must check (P3) and (P4), which we do by finding stochastic bounds for the rate function $\lambda(\boldsymbol{X}(t))$ and for the $\{dN_{ij}(t)\}$ processes. Then, we check existence of the moments for those bounds. In this case, the rate function is

$$\begin{aligned}\lambda(Y(t)) &= \sum_k q_{BY}(Y(t),k) + \sum_k q_{YD}(Y(t),k) \\ &= (\tau^B)^{-1}\ln\left(1+\beta\tau^B Y(t)\right) + (\tau^D)^{-1}\ln\left(1+\delta\tau^D Y(t)\right) \\ &\leq (\beta+\delta)Y(t). \end{aligned} \tag{D23}$$

By (D22) and (D23), it is possible to bound the supremum of the rate function by

$$\bar{\Lambda}_{BY}(t) = \bar{\Lambda}_{YD}(t) \leq (\beta+\delta)\Big(Y(t) + (N_{BY}(t+\bar{h}) - N_{BY}(t))\Big),$$

since the most $\{Y(t)\}$ can be in $[t, t+\bar{h}]$ is the starting value plus all births that occurred during that interval. Similarly, it is possible to bound the supremum of the marginal BY and YD jumps by

$$\bar{Z}_{BY}(t) \leq N_{BY}(t+\bar{h}) - N_{BY}(t) \leq \bar{Z}_{YD}(t) \quad \text{and} \quad \bar{Z}_{YD}(t) \leq Y(t) + \Big(N_{BY}(t+\bar{h}) - N_{BY}(t)\Big),$$

since the birth process can jump at most by all the births occurred in $[t, t+\bar{h}]$ and the death process at most by the starting population plus all those births. Since all these are non-negative,

$$E[\bar{Z}_{ij}(t)\bar{\Lambda}_{ij}(t)|Y(t)=y] \quad \leq \quad E\left[(\beta+\delta)\Big(Y(t) + (N_{BY}(t+\bar{h}) - N_{BY}(t))\Big)^2 |Y(t)=y\right]$$

Then, for both (P3) and (P4) to hold we need to show finiteness of the fourth conditional moment of $N_{BY}(t+\bar{h}) - N_{BY}(t)$. Letting S_{BY} be the event that only transitions of the type BY (and none of type YD) occur in $[t, t+\bar{h}]$,

$$\begin{aligned} E\Big[N_{BY}(t+\bar{h}) - N_{BY}(t)|Y(t)=y\Big] &= \sum kP\Big(N_{BY}(t+\bar{h}) - N_{BY}(t) = k|Y(t)=y\Big) \\ &\leq \sum kP\Big(N_{BY}(t+\bar{h}) - N_{BY}(t) = k|Y(t)=y, \mathbb{I}\{S_{BY}\}=1\Big) < \infty \end{aligned} \tag{D24}$$

where (D24) follows because the probability of k births given that YD transitions have not occurred will not be smaller than otherwise, since those transitions make the birth rate function smaller and move mass in the jump distribution towards zero. To see that (D24) is finite note that the times between transitions, say T_i for $i \geq 1$, occurring in $\{Y(t)\}$ are exponentially distributed with rate $\lambda(y) = (\tau^B)^{-1}\ln(1+\beta\tau^B y) + (\tau^D)^{-1}\ln(1+\delta\tau^D y)$. Now, it is a well known result that T_i has the same distribution as the minimum of $T_{BY,i}$ and $T_{YD,i}$, where these two random variables are independent, exponential random variables with rates $(\tau^B)^{-1}\ln(1+\beta\tau^B y)$ and $(\tau^D)^{-1}\ln(1+\delta\tau^D y)$. Then, conditionally on S_{BY}, (D24) is the expected value of a negative binomial gamma random variable. (P3) and (P4) hold by finiteness of the fourth moment of the negative binomial gamma distribution. □

Note that taking $\tau^B \to 0$ or $\tau^D \to 0$ leads to infinitesimal equi-dispersion in the birth or death process respectively. If both these limits are taken, then $\{Y(t)\}$ becomes a standard, infinitesimally equi-dispersed linear birth-death process. We proceed to check that the models in the SEIR example inherit the dispersion of the building blocks by finding bounds, analogous to those of the proof of Proposition D4. Since our results are for time-homogeneous processes, we consider the time-homogeneous version, i.e. $\nu_{BS}(t) = \nu_{BS}$ and $\beta(t) = \beta$. In this case, letting the whole population at time t be $P(t) = S(t) + I(t) + E(t) + R(t)$, the rate function is

$$\begin{aligned}\lambda(\boldsymbol{X}(t)) &= q_{BS}(\boldsymbol{X}(t),k) + \sum_k q_{SE}(\boldsymbol{X}(t),k) + q_{EI}(\boldsymbol{X}(t),k) + q_{IR}(\boldsymbol{X}(t),k) + \sum_{c\in C} q_{cD}(\boldsymbol{X}(t),k) \\ &= \nu_{BS} + \sum_k q_{SE}(\boldsymbol{X}(t),k) + \nu_{EI}E(t) + \nu_{IR}I(t) + \nu_D P(t). \end{aligned}$$

If the standard simple MCS is considered, $\sum_k q_{SE}(\boldsymbol{X}(t),k) = \beta(I(t)+\gamma)S(t)$. If the compound MCS is considered, $\sum_k q_{SE}(\boldsymbol{X}(t),k) = \tau_{SE}^{-1}\ln\Big(1+\tau_{SE}\beta(I(t)+\gamma)S(t)\Big) \leq \beta(I(t)+\gamma)S(t)$. In both cases,

$$\lambda(\boldsymbol{X}(t)) \leq \nu_{BS} + \Big(\beta + \beta\gamma + \nu_{EI} + \nu_{IR} + \nu_D\Big)P^2(t).$$

Tighter bounds can be derived, but are not needed for our purposes. Bound the supremum of the rate function by

$$\bar{\Lambda}_{ij}(t) \leq \nu_{BS} + \Big(\beta + \beta\gamma + \nu_{EI} + \nu_{IR} + \nu_D\Big)\Big(P(t) + (N_{BS}(t+\bar{h}) - N_{BS}(t))\Big)^2.$$

Similarly, bound the supremum of the marginal jumps by

$$\bar{Z}_{ij}(t) \leq P(t) + (N_{BS}(t+\bar{h}) - N_{BS}(t)).$$

Since all these are non-negative,

$$\begin{aligned} E[\bar{Z}_{ij}(t)\bar{\Lambda}_{ij}(t)|\boldsymbol{X}(t)=\boldsymbol{x}] \quad &\leq \quad E\Big[\nu_{BS}\Big(P(t) + (N_{BS}(t+\bar{h}) - N_{BS}(t))\Big) + \\ &+ \quad \Big(\beta + \beta\gamma + \nu_{EI} + \nu_{IR} + \nu_D\Big)\Big(P(t) + (N_{BS}(t+\bar{h}) - N_{BS}(t))\Big)^3\Big|\boldsymbol{X}(t)=\boldsymbol{x}\Big] \end{aligned}$$

Then, (P3) and (P4) will follow if we establish finiteness of the sixth conditional moment of $N_{BS}(t+\bar{h}) - N_{BS}(t)$. Finiteness follows by

$$E\Big[N_{BS}(t+\bar{h}) - N_{BS}(t)|\boldsymbol{X}(t)=\boldsymbol{x}\Big] \quad = \quad \nu_{BS}\bar{h} < \infty. \tag{D25}$$

To see why (D25) holds, note that, as in the birth-death process above, the inter-event times can be thought of as being exponentially distributed with rate $\lambda(\boldsymbol{x})$ or as the minimum of a collection of independent exponential random variables, each one with rate corresponding to the sum of transition rates regarding a specific transition. In this case, $\{N_{BS}(t)\}$ is a Poisson process with rate ν_{BS} both conditionally and unconditionally on only transitions of the type BS happening in $[t, t+\bar{h}]$. Then, (P3) and (P4) hold by finiteness of the sixth moment of the Poisson distribution. We do not anticipate much difficulty in extending these results to a time-dependent context (e.g. letting ν_{BS} or β depend on time as in Section 4) since the time-varying rates could be replaced in the bounds by their suprema.

[1] Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. J. Royal Stat. Soc. Ser. B Methodol. 72, 269–342.
[2] Applebaum, D., 2004. Lévy Processes and Stochastic Calculus. Cambridge University Press.
[3] Balakrishnan, N., Kozubowski, T., 2008. A class of weighted poisson processes. Stat. & Probab. Lett. 78, 2346–2352.
[4] Bharucha-Reid, A. T., 1960. Elements of the Theory of Markov Processes and their Applications. McGraw-Hill.
[5] Brémaud, P., 1999. Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues. Springer, New York.
[6] Bretó, C., He, D., Ionides, E., King, A., 2009. Time series analysis via mechanistic models. Ann. Appl. Stat. 3, 319–348.
[7] Brown, T. C., Hamza, K., Xia, A., 1998. On the variance to mean ratio for random variables from Markov chains and point processes. J. Appl. Probab. 35 (2), 303–312.
[8] Cox, D., 1955. Some statistical methods connected with series of events. J. Royal Stat. Soc. Ser. B Methodol. 17, 2, 169–174.
[9] Cox, D., Isham, V., 1980. Point Processes. Chapman & Hall.
[10] Daley, D., Vere-Jones, D., 2003. An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods. Springer.
[11] de Quadros, C. A., Andrus, J. K., Danovaro-Holliday, M. C., Castillo-Solórzano, C., 2008. Feasibility of global measles eradication after interruption of transmission in the Americas. Expert Review of Vaccines 7 (3), 355–362.
[12] Economou, A., 2005. Generalized product-form stationary distributions for Markov chains in random environments with queueing applications. Adv. Appl. Probab. 37, 1, 185–211.
[13] Fan, R., Lange, K., Peña, E., 1999. Applications of a formula for the variance function of a stochastic process. Stat. & Probab. Lett. 43, 123–130.
[14] Feller, W., 1971. An Introduction to Probability Theory and Its Applications. Volume II, 2nd Edition. John Wiley & Sons.
[15] Ferrari, M. J., Grais, R. F., Bharti, N., Conlan, A. J. K., Bjornstad, O. N., Wolfson, L. J., Guerin, P. J., Djibo, A., Grenfell, B. T., 2008. The dynamics of measles in sub-Saharan Africa. Nature 451 (7179), 679–684.
[16] Gillespie, C., Renshaw, E., 2005. The evolution of a batch-immigration death process subject to counts. Proc. Royal Soc. Math. Phys. Eng. Sci. 461, 1563 – 1581.
[17] Gillespie, D. T., 1977. Exact stochastic simulation of coupled chemical reactions. Journal of Physical Chemistry 81, 2340–2361.
[18] Gillespie, J., 1984. The molecular clock may be an episodic clock. Proc. National Acad. Sci. United States America 81 No.24, 8009–8013.
[19] Greenwood, M., Yule, G., 1920. An enquiry into the nature of frequency distributions of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. J. Royal Stat. Soc. 83, 255–279.
[20] Haseltine, E. L., Rawlings, J. B., 2002. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. J. Chem. Phys. 117, 6959 – 6969.
[21] He, D., Ionides, E., King, A., 2010. Plug-and-play inference for disease dynamics: Measles in large and small populations as a case study. J. Royal Soc. Interface 7, 271–283.
[22] Hjort, N. L., 1990. Nonparametric Bayes estimators based on beta processes in models for life history data. Ann. Stat. 18, 3, 1259–1294.
[23] Ionides, E. L., Bretó, C., King, A. A., 2006. Inference for nonlinear dynamical systems. Proc. National Acad. Sci. United States America 103, 18438–18443.
[24] Karlin, S., Taylor, H., 1975. A First Course in Stochastic Processes. Academic Press, New York.

[25] Keeling, M., Rohani, P., 2009. Modeling Infectious Diseases in Humans and Animals. Princeton University Press, Princeton, NJ.
[26] King, A., Ionides, E., Pacual, M., Bouna, M., 2008. Inapparent infections and cholera dynamics. Nature 454, 877–880.
[27] Klemm, A., Lindemann, C., Lohmann, M., 2003. Modeling IP traffic using the batch Markovian arrival process. Perform. Evaluation 54, 149 – 173.
[28] Marion, G., Renshaw, E., 2000. Stochastic modelling of environmental variation for biological populations. Theor. Popul. Biol. 57, 197 – 217.
[29] McCullagh, P., Nelder, J. A., 1989. Generalized Linear Models, 2nd Edition. Chapman and Hall, London.
[30] Øksendal, B., 1998. Stochastic Differential Equations, 5th Edition. Springer, New York.
[31] Örmeci, E., Burnetas, A., 2005. Dynamic admission control for loss systems with batch arrivals. Adv. Appl. Probab. 37, 915 – 937.
[32] Ross, S., 1996. Stochastic Processes, 2nd Edition. Wiley.
[33] Sato, K., 1999. Lévy Processes and Inifinitely Divisible Distributions. Cambridge University Press.
[34] Shumway, R. H., Stoffer, D. S., 2006. Time Series Analysis and Its Applications, 2nd Edition. Springer, New York.
[35] Snyder, D., Miller, M., 1991. Randon Point Processes in Time and Space. Springer-Verlag.
[36] Student, 1907. On the error of counting with a haemacytometer. Biometrika 5, 351–360.
[37] Swishchuk, A., Wu, J., 2003. Evolution of Biological Systems in Random Media: Limit Theorems and Stability. Kluwer Acamedic Publishers.
[38] Varughese, M., Fatti, L., 2008. Incorporating environmental stochasticity within a biological population model. Theor. Popul. Biol. 74, 115–129.