Deterministic epidemic models on contact networks: Correlations and unbiological terms

Kieran J. Sharkey

Department of Mathematical Sciences, The University of Liverpool, Peach Street, Liverpool, L69 7ZL, United Kingdom

ARTICLE INFO

ABSTRACT

The relationship between system-level and subsystem-level master equations is investigated and then utilised for a systematic and potentially automated derivation of the hierarchy of moment equations in a susceptible-infectious-removed (SIR) epidemic model. In the context of epidemics on contact networks we use this to show that the approximate nature of some deterministic models such as mean-field and pair-approximation models can be partly understood by the identification of implicit anomalous terms. These terms describe unbiological processes which can be systematically removed up to and including the nth order by nth order moment closure approximations. These terms lead to a detailed understanding of the correlations in network-based epidemic models and contribute to understanding the connection between individual-level epidemic processes and population-level models. The connection with metapopulation models is also discussed. Our analysis is predominantly made at the individual level where the first and second order moment closure models correspond to what we term the individual-based and pair-based deterministic models, respectively. Matlab code is included as supplementary material for solving these models on transmission networks of arbitrary complexity.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Epidemic dynamics are driven by processes which are typically stochastic in nature (Bartlett, 1956; Bailey, 1975). Nevertheless the probabilities of these processes can often be represented or approximated deterministically by a differential form of the Chapman–Kolmogorov equation known as the master equation. This comprehensive set of differential equations describes how the probabilities of the states of a system evolve in time. They are usually too numerous to evaluate numerically although they have been shown to be relevant for small homogeneous epidemic systems (Keeling and Ross, 2008). While numerical solutions remain problematic for systems of any significant size and complexity, master equations do permit exact stochastic realisations using the Gillespie algorithm (Gillespie, 1976; Renshaw, 1991), effectively regenerating the original stochastic epidemic processes.

Master equations are most adept at describing exponentially distributed stochastic processes. This, coupled with the close connection between master equations and other deterministic descriptions of epidemics underlies the almost ubiquitous use of "rates" such as the force of infection and the rate of removal in the design of deterministic epidemic models (Anderson and May, 1991). Although other distributions can be used in principle,

they typically correspond to non-Markovian master equations and this can present significant implementation difficulties. It is worth noting that this is a relatively generic limitation of the deterministic approach.

Several deterministic methods for representing epidemics have been developed. These include the mean-field models (Kermack and McKendrick, 1927; Anderson and May, 1991), pair-approximations (Matsuda et al., 1992; Keeling, 1999; Rand, 1999; van Baalen, 2000; Eames and Keeling, 2002; Murrell et al., 2004; Sharkey et al., 2006), and metapopulation models (Levins, 1969; Sattenspiel and Dietz, 1994; Keeling and Rohani, 2008). All of these attempt to approximate the average time course of an epidemic. Fundamentally, the average time course is implicit in the master equation, but the specific assumptions needed to relate this equation to particular deterministic epidemic models are not always clear. Identification of the relevant assumptions behind these classic deterministic models would certainly enhance our understanding of their domain of applicability and their relationship to the underlying stochastic processes. Indeed, the importance of understanding the basic connection between individual-level processes and population-level deterministic models has been emphasised several times (e.g. Levin and Durrett, 1996; Bansal et al., 2007).

One obvious approach is to start with the master equation and construct solvable deterministic models by applying specific assumptions to directly reduce the dimensionality of the state space. A classic example of this is the Fokker–Planck (or Kolmogorov-forward) equation (Risken, 1989) forming the theoretical justification for reaction–diffusion equation models of epidemics

(Mollison, 1991; Murray, 2003). Another is the van Kampen linear noise approximation to the master equation (van Kampen, 2007) leading to a perturbative volume expansion approach (McKane and Newman, 2004; Ovaskainen and Cornell, 2006). More recently, other more computationally intensive methods of dimensional reduction have been investigated (Sharkey, 2008; Keeling and Ross, 2009; Simon et al., 2010).

In previous work (Sharkey, 2008), a systematic deconstruction of a deterministic SIR epidemic model on arbitrary transmission networks was used to illustrate the connections between the master equation, the network-based mean-field models and the network-based pair-approximation models. Additionally, the individual-based and pair-based deterministic models were constructed. Two assumptions connect these four types of model together with the master equations:- statistical independence and homogeneity. Here we extend this understanding by showing that implicit in the assumption of independence are anomalous terms describing unbiological processes. These terms enable us to understand some of the inaccuracies in network-based deterministic models in a more analytic way than was previously possible.

A side-product of this analysis is a systematic method for obtaining the hierarchy of moment equations at the individual level. The closed form of the first and second order moment equations correspond to the individual-based and pair-based models respectively. Matlab code is provided as supplementary material to solve these models on static network-based systems of arbitrary complexity.

In addition to master equations for systems, master equations for subsections of systems can also be written down. We refer to these subsections as subsystems (Sharkey, 2008). For the present work we start by making a detailed investigation of the relationships between these equations. In particular, we show in the next section that the subsystem master equations follow as a consequence of the system master equation. We also show (Section 3) that conversely, the master equation of a system can be obtained from the master equations of its subsystems provided that the subsystems are statistically independent and, collectively, fully specify the system state. The relevance of this construction for an epidemic system is then briefly introduced.

Section 4 elaborates on the main context for the present work which is a fixed-population susceptible-infectious-removed (SIR) compartmental model on a contact network. It puts this in the context of the general discussion of subsystems and systems, illustrating how moment equations can be derived as a consequence of the system master equation. It also highlights the assumption of pairwise statistical independence which is used to close the first order moment equations. Section 5 discusses the link between this construction and network-based mean-field models and metapopulation models.

Sections 6 and 7 use the results of Section 3 to generate a better understanding of the problems with the pairwise independence assumption. In particular, we show that it generates implicit terms with no obvious interpretation and that these terms allow us to understand the failure of the assumption for certain contact networks. Sections 8 and 9 show how this analysis can be systematically extended beyond the pair level to all orders.

2. Systems, subsystems and master equations

Following prior work (Sharkey, 2008), we start by considering the state Γ^α of an arbitrary system Γ. The probability of the system being in state Γ^α is numerically equivalent to the expectation value $\langle \Gamma^\alpha \rangle$ where here, Γ^α represents a number which has value 1 when the system is in state Γ^α and zero otherwise. We will therefore use $\langle \Gamma^\alpha \rangle$ to denote both the probability of the state

and/or its expectation value. In this notation, the master equation for Γ is:

$$\langle \dot{\Gamma^\alpha} \rangle = \sum_\beta \sigma^{\alpha\beta} \langle \Gamma^\beta \rangle - \sum_\beta \sigma^{\beta\alpha} \langle \Gamma^\alpha \rangle \tag{1}$$

where $\sigma^{\alpha\beta}$ denotes the transition rate from state Γ^β to state Γ^α and here and in what follows, the summations are over all possible system states. Note that this more conventional index ordering is opposite to that used in Sharkey (2008). To avoid any ambiguity, the diagonal elements of all transition matrices in this paper are defined to be zero.

We suppose that within the system Γ, there exist well-defined smaller systems which we refer to as subsystems. We denote these subsystems by ψ_i where the index i distinguishes one subsystem from another. In the next section we will assume that the subsystems do not overlap and that they collectively specify the full system state without ambiguity. Presently we just need to suppose that at least one subsystem of Γ can be identified in an unambiguous manner. We can now write down a set of master equations for the individual subsystem states:

$$\langle \dot{\psi_i^a} \rangle = \sum_b \omega_i^{ab} \langle \psi_i^b \rangle - \sum_b \omega_i^{ba} \langle \psi_i^a \rangle \tag{2}$$

where ω_i^{ab} denotes the transition rate from state ψ_i^b to state ψ_i^a for the subsystem ψ_i. Here and throughout the paper, summations are assumed to be over all of the subsystem states available to ψ_i. We also denote system states by Greek superscripts and subsystem states by Roman superscripts.

Both the system and subsystem master equations must be valid and it is instructive to determine the conditions under which one can be derived as a consequence of the other. Let us attempt to obtain Eq. (2) from Eq. (1). We start with the probability of subsystem ψ_i being in state ψ_i^a which is given by the sum of the probabilities of the system states for which ψ_i is in state ψ_i^a:

$$\langle \psi_i^a \rangle = \sum_\alpha \langle \Gamma^\alpha \rangle D_i^{\alpha a} \tag{3}$$

where $D_i^{\alpha a}$ is a Kronecker-type delta in the states of the subsystems ψ_i such that it has value 1 if the system state Γ^α implies the subsystem state ψ_i^a and zero otherwise:

$$D_i^{\alpha a} = \begin{cases} 1 & \text{if } \Gamma^\alpha \Rightarrow \psi_i^a \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Taking the derivative of Eq. (3) with respect to time and substituting from Eq. (1) gives:

$$\begin{aligned} \langle \dot{\psi_i^a} \rangle &= \sum_\alpha \langle \dot{\Gamma^\alpha} \rangle D_i^{\alpha a} \\ &= \sum_{\alpha\beta} \sigma^{\alpha\beta} \langle \Gamma^\beta \rangle D_i^{\alpha a} - \sum_{\alpha\beta} \sigma^{\beta\alpha} \langle \Gamma^\alpha \rangle D_i^{\alpha a} \\ &= \sum_{\alpha\beta} \sigma^{\alpha\beta} \langle \Gamma^\beta \rangle D_i^{\alpha a} \left[\sum_b D_i^{\beta b} \right] \\ &\quad - \sum_{\alpha\beta} \sigma^{\beta\alpha} \langle \Gamma^\alpha \rangle D_i^{\alpha a} \left[\sum_b D_i^{\beta b} \right] \\ &= \sum_{b\alpha\beta} \sigma^{\alpha\beta} \langle \Gamma^\beta \rangle D_i^{\alpha a} D_i^{\beta b} - \sum_{b\alpha\beta} \sigma^{\beta\alpha} \langle \Gamma^\alpha \rangle D_i^{\alpha a} D_i^{\beta b}. \end{aligned} \tag{5}$$

Swapping the dummy indices α and β in the second term on the right gives:

$$\langle \dot{\psi_i^a} \rangle = \sum_{b\alpha\beta} \sigma^{\alpha\beta} \langle \Gamma^\beta \rangle D_i^{\alpha a} D_i^{\beta b} - \sum_{b\alpha\beta} \sigma^{\alpha\beta} \langle \Gamma^\beta \rangle D_i^{\beta a} D_i^{\alpha b}. \tag{6}$$