PAIR APPROXIMATIONS

JOSHUA L. PAYNE
Dartmouth College, Lebanon, New Hampshire

Pair approximations are analytical techniques for estimating the dynamics and equilibrium properties of network-based models. As the name implies, pair approximations capture the dynamics of the states of neighboring pairs of vertices in a network, as opposed to the dynamics of individual vertex states. These methods have been successfully applied to a variety of network-based ecological and evolutionary models, ranging from the evolution of cooperation to the spread of infectious disease.

MODELING INTERACTIONS

Many ecological processes occur on spatial scales that are much smaller than the entire geographic range of a population. To model such local interactions, populations are often represented as networks, where vertices denote individuals and edges denote their interactions. Finding exact analytical solutions to models of dynamical processes on networks is often exceedingly difficult. Pair approximations are commonly used to overcome these difficulties. Instead of providing an exact solution, these methods use differential equations to approximate the rates of change in the states of connected pairs of vertices, allowing for an estimation of the model's dynamics and equilibrium properties.

THE *SIS* MODEL OF DISEASE SPREAD

Pair approximations will be presented in the context of the classical Susceptible–Infected–Susceptible (*SIS*) model of disease spread, in order to provide a concrete example of their application. In the *SIS* model, a population of N individuals is compartmentalized into two discrete states: susceptible (S) and infected (I). A susceptible individual does not have the disease but is vulnerable to it, whereas an infected individual has the disease and the potential to pass it on.

In the original formulation of this model, which is commonly referred to as a mass-action model, individuals are assumed to come in contact with one another randomly, at rate β individuals per unit time. Letting $[S]$ and $[I]$ denote the number of susceptible and infected individuals in the population, the dynamics of disease spread ($\frac{d[I]}{dt}$) can be simply described with the following differential equation

$$\frac{d[I]}{dt} = \beta\frac{[S]}{N}[I] - g[I]. \quad (1)$$

The rate of change in the number of infected individuals reflects a balance between infection and recovery events. $[I]$ increases at rate $\beta\frac{[S]}{N}[I]$, because each of the $[I]$ infected individuals come into contact with β individuals per unit time, of which $\frac{[S]}{N}$ are susceptible. $[I]$ decreases at rate $g[I]$, as infected individuals recover and return to the susceptible state. The ratio between the contact rate β and the recovery rate g is known as the basic reproductive ratio $R_0 = \beta/g$, which determines whether or not a disease will spread throughout a population. Specifically, if $R_0 > 1$ the contagion will spread because each infected individual transmits the disease to, on average, more than one susceptible individual.

PAIR APPROXIMATIONS

The assumption that individuals encounter one another at random is an oversimplification of the interaction patterns of natural systems; individuals typically encounter

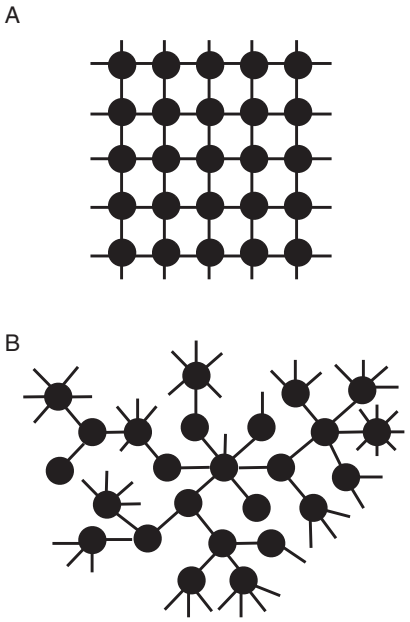

FIGURE 1 Two examples of commonly used interaction networks. (A) Lattice interaction network where each vertex is connected to its nearest neighbors. (B) Random interaction network, with the same average number of edges per vertex as in (A). For illustration, only a small portion ($N = 25$) of the entire network is depicted. Dangling edges denote connections to vertices that are not shown.

only a small fraction of the total population in their lifetime, and these interactions are usually not random. Such structured interactions are often captured using networks, where individuals are represented as vertices and inter-individual interactions are represented as edges (Fig. 1). When a disease spreads throughout such a structured population, its dynamics deviate from those provided by Equation 1, because the assumption of random interaction is violated. The probability of an individual contracting the disease depends on whether or not any of its neighbors in the network are infected. Thus, correlations exist between the states of connected pairs of vertices.

Pair approximations use differential equations to explicitly track these correlations. They were first applied to epidemiological models by Matt Keeling (1999), and we will use his approach and notation. Consider a population structured on a network where every vertex has k edges, and disease transmissibility across an edge is given by $\tau = \beta/k$. Let $[SI]$ denote the number of pairs of connected vertices where one vertex is susceptible and the other is infected (the terms $[SS]$ and $[II]$ are similarly defined, but are counted twice; i.e., these quantities are always even), and let $[SSI]$ denote any connected three-vertex configuration (referred to as a triplet) where the first two vertices are susceptible and the last is infected ($[ISI]$ is similarly defined).

To estimate the dynamics of disease spread, we need to monitor the coupled dynamics of four quantities: $[SS]$, $[SI]$, $[IS]$, and $[II]$. However, we can exploit both symmetry ($[SI] = [IS]$) and redundancy ($[II] = Nk - [SS] - 2[SI]$) so that we only have to monitor the rate of change in $[SI]$ and $[SS]$. The quantity $[SI]$ can change in five ways: It can increase if

(i) one of the vertices in an II pair reverts back to the susceptible state ($II \Rightarrow SI$), which occurs at rate g, or

(ii) one of the vertices in an SS pair contracts the disease from an infected individual outside the pair ($SS \Rightarrow SI$), which occurs at rate τ.

It can decrease if

(iii) an infected vertex in an SI pair reverts back to the susceptible state ($SI \Rightarrow SS$), which occurs at rate g,

(iv) an infected vertex in an SI pair transmits the disease to the susceptible vertex ($SI \Rightarrow II$), which occurs at rate τ, or

(v) a susceptible individual in an SI pair contracts the disease from an infected individual outside the pair ($SI \Rightarrow II$), which occurs at rate τ.

Conditions (ii) and (v) both require information about a vertex state outside of the SI pair. Specifically, condition (ii) can only occur if an SS pair is part of an SSI triplet, and condition (v) can only occur if an SI pair is part of an ISI triplet.

Thus, the rates of change in the number of pairs depend upon the numbers of configurations larger than pairs, and this information is not available. These higher-order quantities are approximated by assuming that the vertices at the opposing ends of a triplet are independent of one another (i.e., triplets form linear chains, not triangles). Under this assumption, $[SSI]$ can be approximated as

$$[SSI] = \frac{(k-1)\,[SS][SI]}{\Sigma_{X \in S,I}\,[SX]} = \frac{(k-1)}{k}\,\frac{[SS][SI]}{[S]}, \quad (2)$$

where the last equality is valid because the number of singles (e.g., $[S]$) can always be recovered from the number of pairs,

$$[S] = \frac{1}{k}\sum_{X \in S,I}[SX]. \quad (3)$$

The approximation of higher-level quantities from their lower-level counterparts is referred to as "closing" the system, and the name "pair approximation" comes from the fact that this system is closed at the level of pairs.

Using Equations 2 and 3 and the transition rules (i)–(v), we can now describe the rate of change in $[SI]$ as

$$\frac{d[SI]}{dt} = \overbrace{g[II]}^{(i)} + \overbrace{\tau[SSI]}^{(ii)} - \underbrace{g[SI]}_{(iii)} - \underbrace{\tau[SI]}_{(iv)} - \underbrace{\tau[ISI]}_{(v)}. \quad (4)$$

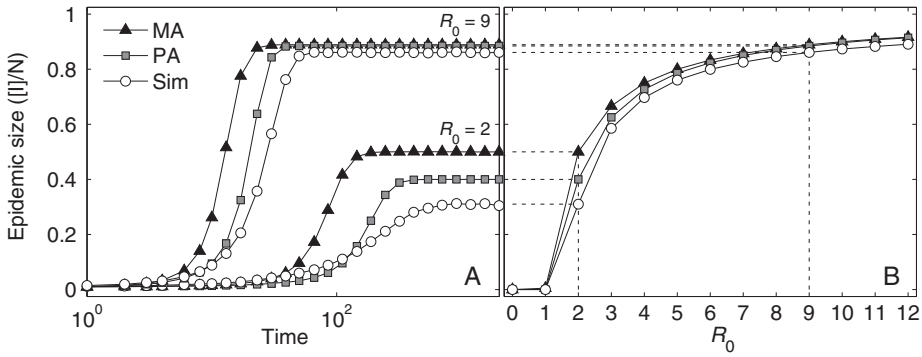

FIGURE 2 (A) Dynamics and (B) equilibrium size of epidemic outbreaks as estimated by the mass-action model (MA, Eq. 1), the pair approximation (PA, Eqs. 4 and 5), and as observed via direct simulation (Sim) on a 10×10 square lattice with nearest-neighbor interactions ($k = 4$). Simulation results correspond to 1000 independent replications for each value of the reproductive ratio $R_0 = \beta/g$. In (A), the x-axis is logarithmically scaled.

Similarly, we can describe the rate of change in $[SS]$ as

$$\frac{d[SS]}{dt} = 2g[SI] - 2\tau[SSI], \qquad (5)$$

where the first term on the right-hand side of the equation captures the infected individuals in SI pairs reverting back to susceptibility, and the second term captures SS pairs changing to SI pairs due to their involvement in SSI triplets. These quantities are doubled to ensure that $[SS]$ is counted twice, which is required by Equation 3.

COMPARING THEORY AND DATA

In Figure 2, we depict the dynamics and equilibrium conditions of disease spread estimated by the mass-action model (Eq. 1) and the pair approximation (Eqs. 4 and 5), and as observed through direct simulation on a square lattice with nearest-neighbor interactions (Fig. 1A). As expected, the mass-action model consistently overpredicts the rate of disease spread, relative to the simulations (Fig. 2A). The pair approximation offers considerable improvement, estimating a slower rate of spread than the mass-action model, though not quite as slow as that observed through simulation.

The dynamics estimated by the pair approximation are therefore not completely accurate. However, the equilibrium outbreak size is generally well captured, and for low R_0, the pair approximation provides a more accurate estimation of equilibrium outbreak size than the mass-action model (Fig. 2B).

DISCUSSION

By tracking the correlations between the states of connected vertices, pair approximations provide a more accurate description of the dynamics and equilibrium conditions of disease spread through structured populations than the mass-action model. The discrepancies observed between the pair approximation and the simulation data result from the violation of one principal assumption of the closure method used in the pair approximation: that the underlying contact network is perfectly branching (i.e., possesses no loops, as in Fig. 1B). In that case, it is accurate to assume that the distant ends of triplets are completely independent of one another, as is done in Equation 2. However, in the lattice network considered herein (Fig. 1A), loops abound, and they considerably impact the rate of disease spread.

The accuracy of the pair approximation in predicting the pre-equilibrium dynamics of disease outbreaks can be improved considerably by taking into account some of the topological features of the network. For example, the proportion of triplets that form closed triangles can be incorporated into the closure method (Eq. 2), or the ratio of the local neighborhood size to the underlying lattice size can be used to parameterize the differential equations that describe the rate of disease spread (Eqs. 4 and 5). Accuracy can also be improved by explicitly tracking the dynamics of higher-order motifs, such as triplets. However, the required number of differential equations grows exponentially with the size of the motifs being tracked, which is why the system is usually closed at the level of pairs.

AREAS OF APPLICATION

The pair approximation is a versatile technique that has been applied to many network-based models of ecological and evolutionary processes. For example, it has been used to derive explicit conditions for species invasions in viscous populations, particularly in the context of vegetation dynamics. The pair approximation has also been used to derive simple rules for the evolution of cooperative behavior in social dilemmas and to estimate the equilibrium proportion of cooperators in various evolutionary games. As discussed in this entry, the pair

approximation has also been applied to epidemiology, capturing the rate of disease spread and the final epidemic size in structured populations.

SEE ALSO THE FOLLOWING ARTICLES

Adaptive Dynamics / Cooperation, Evolution of / Disease Dynamics / Epidemiology and Epidemic Modeling / Networks, Ecological

FURTHER READING

Keeling, M. J. 1999. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society London B: Biological Sciences* 266: 859–8767.

Keeling, M. J., and K. T. D. Eames. 2005. Networks and epidemic models. *Journal of the Royal Society Interface* 2: 295–307.

Newman, M. E. J. 2010. *Networks: an introduction.* Oxford: Oxford University Press.

Payne, J. L., and M. J. Eppstein. 2009. Pair approximations of takeover dynamics in regular population structures. *Evolutionary Computation* 17: 203–229.

Petermann, T., and P. De Los Rios. 2004. Cluster approximations for epidemic processes: a systematic description of correlations beyond the pair level. *Journal of Theoretical Biology* 229: 1–11.

Satō, K., and Y. Iwasa. 2000. Pair approximations for lattice-based ecological models. In U. Dieckmann, R. Law, and J. A. J. Metz, eds. *The geometry of ecological interactions: simplifying spatial complexity*. Cambridge, UK: Cambridge University Press.

van Baalen, M. 2000. Pair approximations for different spatial geometries. In U. Dieckmann, R. Law, and J. A. J. Metz, eds. *The geometry of ecological interactions: simplifying spatial complexity*. Cambridge, UK: Cambridge University Press.

PARTIAL DIFFERENTIAL EQUATIONS

NICHOLAS F. BRITTON

University of Bath, United Kingdom

There are extensive applications of partial differential equations (PDEs) in ecology, covering all the main aspects of the science. Questions about the distribution and abundance of organisms may involve PDEs for the population density of various species depending on time and space, for example, in a process of ecological succession. Questions about the movement of materials and energy through living communities may involve PDEs for the concentration of particular chemicals, again depending on time and space. The age structure of a population may change over time according to a PDE, and it may be crucial in considering life processes. The genotypic structure of a population or of a community, which may with time move through some trait space according to a PDE, may be analyzed to explain adaptations and coadaptations. PDEs in science, and in ecology in particular, are often mathematical expressions of conservation laws (such as the law of conservation of mass) and are therefore based on a sound conceptual foundation. Modeling a single population in isolation will generally lead to a single PDE, while models of interacting populations or the interaction of a population with an abiotic resource will lead to systems of coupled PDEs.

SCOPE OF PDE MODELS

A PDE model is not appropriate unless the independent variables are continuous. Time is essentially continuous, but an insect population with nonoverlapping generations is often modeled in discrete time, and although space is essentially continuous, a population in a patchy habitat may be modeled as occupying discrete space. Insects pass through discrete stages in their life history, in a model of a disease the host population is often divided into a finite number of classes, and the simplest population-genetic models consider a finite number of genotypes. Such populations cannot be described using PDEs in the time, space, or structure variables, respectively.

A PDE model is also inappropriate if the dependent variables are not continuous functions of the independent variables, or at least if they may not be approximated as continuous functions. For example, phosphate uptake by phytoplankton in the ocean depends on the phosphate concentration u, which varies as a function of space $\mathbf{x} = (x, y, z)$ and time t. Concentration may be defined as amount of substance per unit volume. The concentration of phosphate in a volume of water containing a point $\mathbf{x}$ is therefore defined, but its concentration *at* the point $\mathbf{x}$ is not. We circumvent this problem by using a continuum approximation: we consider the concentration of phosphate at a point to be its concentration in a volume containing the point where the volume is small compared to the phytoplankton, but large enough that the discrete nature of phosphate ions does not have to be taken into account. This separation in spatial scales is crucial to the approximation. We may define the population density of the phytoplankton similarly, either as the number of individual phytoplankton or as the biomass of phytoplankton per unit volume of water, but it may be that the nature of phytoplankton as discrete organisms plays an essential role in determining the behavior under investigation. If this is so, a PDE model is inappropriate and an individual-based model is required.

In this article we shall consider PDE models for biological populations and abiotic resources distributed in