On a combinatorial problem in botanical epidemiology

Toufik Mansour [a,*], Matthias Schork [b]

[a] Department of Mathematics, University of Haifa, 3498838 Haifa, Israel
[b] Camillo-Sitte-Weg 25, 60488 Frankfurt, Germany

ARTICLE INFO

ABSTRACT

In botanical epidemiology, one of the major problems is to study the spread of diseases within crops. Several approaches to study the patterns and temporal evolution of diseases have been discussed in the literature, e.g., statistical techniques or variogram analysis. Recently, AlSharawi, Burstein, Deadman and Umar determined the total number of ways to have ℓ isolated infected individuals among m infected plants in a row of n plants. They discussed this in a straightforward combinatorial fashion and considered several associated points, like expectation value and variance of the number of isolated infected plants. In the present paper, we derive their results with the help of generating function techniques and use this method to extend the discussion to plants arranged in a circle as well as in two rows. The dependence of the expected number of isolated infected plants on the proportion of infected plants is considered for large n in all these cases as well as in the case of an arbitrary number of rows and a simple asymptotic behavior is found. Connections to several combinatorial sequences are established.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Plant epidemiology is, by definition, concerned with the study of populations of pathogens in populations of plants. In particular, it is the aim to understand the relationship between epidemic progress (i.e., spatial and temporal development), distribution of disease in the field and pathogen dispersal. Due to its complex interdisciplinary character and the multiple scales in space and time involved, progress is slow, although the knowledge gained is of great ecological and economical importance. As a general reference for plant epidemiology, we refer to [4]. Furthermore, we would like to point out the reviews [8,13,21,25] closer to the contents of the present paper. When considering plant epidemiology, i.e., the interaction between a plant (or a host) and a pathogen, one has the advantage that the plants cannot move, in contrast to interactions between populations of animals, e.g., in predator–prey systems. In agriculture, plants are often arranged in either a homogeneous mixture or in a regular grid lattice. As discussed in [19], for such configurations the pathogen distributions can be analyzed by looking at the spatial pattern of disease in absolute spatial coordinates. In natural systems, the plant population itself may show a spatial pattern over which the disease process is superimposed. Thus, disease may be spatially aggregated purely due to the spatial aggregation of the underlying plant population, and in any analysis of disease pattern in plants it will be useful to factor out the underlying spatial plant distribution so as to avoid confusion over the true pattern of disease occurrence within its host.

However, even for the simpler case of agriculture where the plants are assumed to be distributed regularly in space, there are several scales involved. For the plant, one can consider as unit of consideration, e.g., the plant as a whole, or its

* Corresponding author.
  E-mail addresses: tmansour@univ.haifa.ac.il (T. Mansour), mschork@member.ams.org (M. Schork).

leaves, or its cells, and for the infection one can consider the dynamics from cell to cell, plant to plant or field to field (with natural limit the entire continent). In addition to these different spatial scales, there are several time scales involved, arising, e.g., from the reproduction rate of the pathogen versus the one of the host, the natural time scales involved (e.g., seasons) or the life-time of the plant (e.g., crop versus trees).

For a given plant–pathogen system under consideration, one may ask the following "direct" questions: Suppose the mechanism of infection is known, how does the infection develop in time and space? Which patterns do arise? Will a stable distribution of pathogens be reached? If not, what will happen instead? If the infection is lethal, will there be an extinction of the host? How can the development be controlled or managed to obtain a desired outcome? What will happen when several pathogens are involved which may interact in a direct or indirect fashion (e.g., via cross protection [26])? It is also possible to ask "inverse" questions: Suppose patterns of infected plants are observed, what can be inferred about the mechanism of infection? Or the locations of the origins? If a sample is drawn, to what degree can data and model be transferred to another (smaller or larger) scale [16]? To be able to answer these questions, one has to consider appropriate models. As mentioned above, there are several scales involved, leaving many degrees of freedom when setting up a model. For example, if one considers a disease that spreads from plant to plant, one has to decide how the plants are distributed in space (e.g., in a regular fashion on a lattice [20], or in an arbitrary network [14]), how the neighbors of a plant are defined (e.g., nearest neighborhood connectedness on a lattice, Gabriel connectedness, via some metric on the surrounding space) and how the mechanism of infection works. The choice of a connectedness scheme depends on the specific questions and assumptions about the biological system, see, e.g., [19] for a discussion of some often used schemes. If one considers several spatial hierarchies (scales) at once (e.g., field, plant, leaf and cell), there are further parameters involved, describing the interaction between the different hierarchies, see, e.g., [9].

To understand the dynamics of an infection and the resulting patterns of infected plants, many approaches have been considered, see, e.g., [2,3,5,6,10,14,17–20,22,23]. In particular, in [17,19], several approaches are discussed to how patterns of diseased plants can be characterized with the help of different statistical measures. The conclusion of these works is that a suite of methods for the description and analysis of point patterns in plant ecology or epidemiology should be used since no single descriptor is able to capture all relevant aspects in a satisfactory way.

In most of the models discussed, the disease progresses from isolated foci which expand via neighboring plants. In a recent paper, AlSharawi, Burstein, Deadman and Umar [1] argued that one should consider amenity plantings and certain row crops which would be essentially one dimensional with no interference between rows. Therefore, significant information could be gained by analyzing the extent of clustering of infected individuals along rows, or, conversely, the extent to which infected individuals are isolated along rows. They showed that the total number of configurations having ℓ isolated infected individuals among m infected plants in a row of n plants (or cells) is given for $n \geq m \geq \ell \geq 0$ such that $m \geq \ell + 2$ by

$$F(n, m, \ell) = \binom{n - m + 1}{\ell} \sum_{j \geq 0} \binom{n - m - \ell + 1}{j + 1} \binom{m - \ell - j - 2}{j}. \tag{1.1}$$

Furthermore, for fixed n and m, they determined the expected number of isolated infected plants and its variance. The methods they used consisted of a straightforward – although slightly tedious – combinatorial enumeration and a beautiful stochastic interpretation. The aim of the present paper is twofold: On the one hand, we want to advertise the use of generating functions as a powerful tool for these kind of questions. On the other hand, we want to extend the results of [1] to a more realistic setting using these methods, the ultimate goal being the description of an arbitrary field consisting of r rows and n columns. Along the way, connections to several combinatorial sequences are established.

The structure of the paper is as follows. In Section 2, we consider a single row of n plants of which m are infected and consider the configurations where ℓ of the infected plants are isolated infected. We derive several of the results of AlSharawi, Burstein, Deadman and Umar [1] using generating functions and consider the dependence of the expectation value of isolated infected plants on n and the proportion α of infected plants for large n. In Section 3, plants arranged in a circle will be considered in the same fashion. In Section 4, we extend this treatment to plants arranged in two rows (each having n plants) and derive explicit and asymptotic results for the number of configurations where in total m plants are infected of which ℓ are isolated infected. Some of these results are extended to the case of an arbitrary number of rows of plants in Section 5. Finally, in Section 6, some conclusions are drawn.

2. A single row of plants

Let us assume there are n plants in a single row such that any plant can be either infected by one of the $k - 1$ diseases or not be infected; here $k \geq 2$ is a natural number. Each possible arrangement of such plants can be represented as a word $\pi_1 \pi_2 \cdots \pi_n$ of length n such that $\pi_i \in [k - 1] = \{0, 1, \ldots, k - 1\}$, where $\pi_i = 0$ when the ith plant is not infected and $\pi_i = j$ with $1 \leq j \leq k - 1$ when the ith plant is infected by the jth disease. Let $R_{n|m}^{(1)}$ be the set of all words – or *configurations* – where m plants are infected by one of the diseases $1, \ldots, k - 1$ in a row of n plants. The superscript "(1)" is used to denote that we consider one row (in later sections, we will consider two and more rows). A plant π_i will be called *isolated infected* if $1 \leq \pi_i \leq k - 1$, i.e., it is infected, and $\pi_{i-1} = 0 = \pi_{i+1}$, i.e., both of its neighbors are not infected (here we set formally $\pi_{-1} = 0 = \pi_{n+1}$). We denote the generating function for the number of configurations having ℓ isolated infected individuals