



Using Optimal Transport Theory to Estimate Transition Probabilities in Metapopulation Dynamics



J.M. Nichols^{a,*}, J.A. Spendelov^b, J.D. Nichols^b

^a Naval Research Laboratory, 4555 Overlook Ave., Washington, D.C. 20375, United States

^b U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, MD 20708, United States

ARTICLE INFO

Article history:

Received 3 March 2017

Received in revised form 6 June 2017

Accepted 7 June 2017

Keywords:

Lagrangian
Optimal transport
Population dynamics
Transition probabilities
Movement ecology
Cost-energy function

ABSTRACT

This work considers the estimation of transition probabilities associated with populations moving among multiple spatial locations based on numbers of individuals at each location at two points in time. The problem is generally underdetermined as there exists an extremely large number of ways in which individuals can move from one set of locations to another. A unique solution therefore requires a constraint. The theory of optimal transport provides such a constraint in the form of a cost function, to be minimized in expectation over the space of possible transition matrices. We demonstrate the optimal transport approach on marked bird data and compare to the probabilities obtained via maximum likelihood estimation based on marked individuals. It is shown that by choosing the squared Euclidean distance as the cost, the estimated transition probabilities compare favorably to those obtained via maximum likelihood with marked individuals. Other implications of this cost are discussed, including the ability to accurately interpolate the population's spatial distribution at unobserved points in time and the more general relationship between the cost and minimum transport energy.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Movement of individuals from one subpopulation to another is a key determinant of population dynamics and a cornerstone of metapopulation theory (Hanski, 1998, 1999; Turchin, 1998). In animal ecology, inferences about rates and underlying probabilities of movement are typically based on studies of marked individuals. In some cases the marks are radio transmitters, and animal location can be determined very frequently and at will (White and Garrott, 1990; Patterson et al., 2007). For other individual marks, animal location cannot be determined remotely, and must be assessed via recapturing or resighting marked individuals. These sampling methods admit nondetection of marked animals that are present in sampled areas, and require multistate capture-recapture models (e.g., Arnason, 1972, 1973; Brownie et al., 1993; Schwarz et al., 1993; Lebreton et al., 2009) for inference about movement. Both types of study require substantial effort and expense.

An alternative approach to inference about movement can be based on aggregate data in which the identity of individual animals is not available (e.g., Willekens, 1977; Willekens et al., 1981; Cooch and Link, 1999). For example, consider a certain number of

individual animals that are present in our system at both times t and $t + 1$. We do not know the identities of each individual at each time period, but we instead have counts of how many of them are at each of N locations at both times. Depending on the number of individuals involved, there can be a large number of possible individual movements that produce any set of location-specific counts at $t + 1$, given the counts at t (Gail and Mantel, 1977). Hence, inferences about actual rates of movement are not possible without constraints on the problem. Cooch and Link (1999) used an entropy maximization approach to inference but concluded that the requisite assumptions were sufficiently restrictive to limit utility of the approach for ecological systems.

Here we present a different approach to this problem of inference about rates of movement using only aggregate data. The approach derives from the theory of optimal transport (Villani, 2008), in which the requisite constraint takes the form of a cost function to be minimized. Although the development of optimal transport theory has taken place in disciplines other than ecology, its underlying conceptual framework is closely related to ideas in metapopulation theory and landscape ecology. For example, ecologists are frequently concerned about fragmented landscapes that reduce animal movement, using the term “connectivity” to refer to “the degree to which the landscape facilitates or impedes movement of organisms among source patches” (e.g., Taylor et al., 1993; Tischendorf and Fahrig, 2000). Connectivity is typically expressed

* Corresponding author.

as some function of the cost required to move between patches, and the cost is usually either squared Euclidean distance or some function of it (Moilanen and Hanski, 2001; Adriaensen et al., 2003). In addition, the very theory of optimal transport is based on the underlying principle of work minimization in physical processes, whereas the process of natural selection tends to maximize fitness and associated behaviors of biological organisms (e.g., Fisher, 1930; Clark and Mangel, 2000).

In what follows we provide a brief description of optimal transport theory as it pertains to the problem at hand: namely the estimation of transition probabilities governing movement based only on aggregate summary statistics at two different times. Section 2 describes the basic optimal transport problem, while Section 3 discusses the cost of movement and how that cost is related to system dynamics. Section 4 then provides an illustrative numerical example, and Section 5 applies the theory to a real-world example, comparing the method to a more conventional approach based on maximum likelihood estimation with marked individuals. We discuss the implications of the analysis in Section 6.

2. Problem formulation

The basic optimal transport problem is to minimize the expected cost $E_{XY}[c(\mathbf{x}, \mathbf{y})]$ of moving a unit “mass” from one spatial location, \mathbf{x} , to another, \mathbf{y} . The distribution of mass at the starting and ending locations is characterized by the probability functions $P_X(\mathbf{x})$ and $P_Y(\mathbf{y})$; these are considered known at the problem outset (In our notation we use capital letters, e.g., X , to denote random variables and lower case, e.g., \mathbf{x} , to denote the values they are used to model). Moreover, the random variables X, Y will be defined on the problem domain Ω , taken here as the subset of \mathbb{R}^2 describing the particular patches of earth on which our populations are distributed.

What is unknown is the joint probability distribution $P_{XY}(\mathbf{x}, \mathbf{y})$, defined on $\Omega \times \Omega$, describing the fraction of mass at each location in \mathbf{x} that must move to each location in \mathbf{y} . This distribution is referred to as the “optimal transport plan” in the sense that it minimizes expected cost via

$$K(P_X(x), P_Y(y)) = \min_{P_{XY}(\mathbf{x}, \mathbf{y})} \int_{\mathbf{x} \times \mathbf{y}} c(\mathbf{x}, \mathbf{y}) dP_{XY}(\mathbf{x}, \mathbf{y}). \quad (1)$$

subject to the constraint that $P_{XY}(\mathbf{x}, \mathbf{y})$ admits $P_X(\mathbf{x})$ and $P_Y(\mathbf{y})$ as marginals.

Eq. (1) is known in optimal transport theory as the Kantorovich distance. Remarkably, (1) has a unique minimizer despite the infinite number of possible transport plans that are consistent with both $P_X(\mathbf{x})$ and $P_Y(\mathbf{y})$ as marginals. The only practical requirement is that the cost function return a positive, real valued number (Villani, 2008). While the roots of the minimization problem (1) extend back a hundred or more years (see again Villani (2008) or Bogachev and Kolesnikov (2012) for historical treatment), a tremendous volume of recent work (last 10–15 years) has explored the theoretical and computational aspects of this minimization problem.

In an ecological context, one can think of $P_X(\mathbf{x})$ and $P_Y(\mathbf{y})$ as population distributions at times t and $t+T$ respectively while each location in Ω is specified by two coordinates, $\mathbf{x}=(x_1, x_2)$, $\mathbf{y}=(y_1, y_2)$, defining a position on the earth’s surface relative to a user-defined origin. We will correspondingly denote as N and M the number of such locations associated with the starting and ending distributions respectively. The i th such location will be denoted $\mathbf{x}_i \equiv (x_{1,i}, x_{2,i})$. The distributions we consider are discrete, e.g., $P_X(\mathbf{x}_i) = \int_X \frac{B_X(i)}{B_T} \delta_X(\mathbf{x} - \mathbf{x}_i) d\mathbf{x}$, $i = 1 \dots N$ and $P_Y(\mathbf{y}_j) = \int_Y \frac{B_Y(j)}{B_T} \delta_Y(\mathbf{y} - \mathbf{y}_j) d\mathbf{y}$, $j = 1 \dots M$, where $\delta(\cdot)$ is the Dirac delta function. Here we will use B_T to denote the total number of individuals present at both times, t and $t+T$, while $B_X(i)$, $i = 1 \dots N$, $B_Y(j)$,

$j = 1 \dots M$ represent population counts at spatial sites $\mathbf{x}_i, \mathbf{y}_j$ respectively.

The discrete, joint distribution for which we are solving, $P_{XY}(\mathbf{x}_i, \mathbf{y}_j)$, tells us how many individuals at location \mathbf{x}_i at time t move to location \mathbf{y}_j at time $t+T$. The discrete version of (1) is therefore given by Kolouri et al. (2016)

$$\begin{aligned} K(P_X(\mathbf{x}_i), P_Y(\mathbf{y}_j)) &= \min_{P_{XY}(\mathbf{x}_i, \mathbf{y}_j)} \sum_{i=1}^N \sum_{j=1}^M c(\mathbf{x}_i, \mathbf{y}_j) P_{XY}(\mathbf{x}_i, \mathbf{y}_j) \\ \text{s.t. } \sum_{j=1}^M P_{XY}(\mathbf{x}_i, \mathbf{y}_j) &= P_X(\mathbf{x}_i), \sum_{i=1}^N P_{XY}(\mathbf{x}_i, \mathbf{y}_j) = P_Y(\mathbf{y}_j) \\ P_{XY}(\mathbf{x}_i, \mathbf{y}_j) &\geq 0, i = 1 \dots N, j = 1 \dots M \end{aligned} \quad (2)$$

and can be used to solve for the $N \times M$ matrix of transport probabilities, $P_{XY}(\mathbf{x}_i, \mathbf{y}_j)$, $i = 1 \dots N, j = 1 \dots M$.

This information is typically estimated by tracking a representative subset of individuals in the population from t to $t+T$. The minimizer (2) affords the possibility of obtaining this same information, but without tracking individuals; rather the desired probabilities are obtained by supplying the appropriate cost function with marginals as constraints and solving (2). Note, that (2) is a linear program (albeit a potentially high-dimensional one) and can be solved using standard numerical methods.

While the number and locations of the monitoring sites can be different at times t and $t+T$, in what follows we will assume a fixed number of monitoring sites so that $M=N$ and $\mathbf{y}_i = \mathbf{x}_i, i = 1 \dots N$. This is likely to be the more typical situation in practice (Spendlow et al., 1995; Martin et al., 2006; Sanderlin et al., 2012). See Fig. 3, Section 5 for a graphical picture of the relevant quantities and associated notation for our example study system.

3. Choice of cost

Defining an appropriate cost of movement has been discussed previously in the ecological literature, and may be a direct function of Euclidean distance or instead modified by features such as elevational gradient, habitat, etc. (Adriaensen et al., 2003; Bonte et al., 2012; Etherington, 2016). Frequently, the cost is taken as simply the distance over which members of the population must travel, i.e., longer distances equate with higher cost. In fact, we will show that by choosing $c(\mathbf{x}, \mathbf{y}) = \|\mathbf{y} - \mathbf{x}\|_2^2$, the solution to (1) produces a transport plan that minimizes the kinetic energy associated with movement. In doing so we will review the well-defined connection between “cost” and “energy” of optimal transport. In our view, this relationship underscores the potential power of optimal transport in studying metapopulation dynamics.

3.1. Cost as an energy minimizer

In mechanics, the principle of energy minimization guides the derivation of equations governing the dynamics of many types of systems (e.g., structural dynamics, fluid mechanics, thermodynamics, etc.). In short, the principle states that of all possible paths describing the evolution of a dynamical system, nature will take the one associated with the least energy (least work).

Denote the time-dependent state of a system by the vector $\mathbf{y}_t \in \mathbb{R}^d$ and its time derivative $\dot{\mathbf{y}}_t$. The Lagrangian of a dynamical system, denoted $L(\mathbf{y}_t, \dot{\mathbf{y}}_t, t)$, quantifies the work associated with non-dissipative (conservative) forces and can be written as the difference between the system kinetic and potential energy (Nichols and Murphy, 2016). Solutions \mathbf{y}_t that minimize the total

Download English Version:

<https://daneshyari.com/en/article/5742143>

Download Persian Version:

<https://daneshyari.com/article/5742143>

[Daneshyari.com](https://daneshyari.com)