1 Discrete-Time Transition Models

1.1 Definitions

Definition 1.1. Transition Model A *transition model* is a quintuple $\Theta = (V, K, T, \boldsymbol{\rho})$ consisting of the following:

1. V, a set of nodes $V = \{1, 2, \ldots, n\}$
2. K, a set of discrete times $K = \{0, 1, \ldots, L\}$
3. T, a *transition function* $T : K \to \mathrm{M}_n$
4. $\boldsymbol{\rho}$, a *population density function* $\boldsymbol{\rho} : K \to \mathbb{R}^n$

These elements obey the following properties:

1. For each $k \in K - \{L\}$, $T(k)$ is defined, and $T(k)$ is a right stochastic $n \times n$ matrix.
2. For each $k \in K$, $\boldsymbol{\rho}(k)$ is defined, and $\boldsymbol{\rho}(k)$ is a stochastic column vector.
3. For each $k \in K - \{L\}$, it is true that $\boldsymbol{\rho}(k+1) = T(k)\boldsymbol{\rho}(k)$.

Definition 1.2. Individuals Let $\Theta = (V, K, T, \boldsymbol{\rho})$ be a transition model. An *individual* ι on Θ is a mapping $\iota : K \to V$, such that $\iota(k)$ is defined for all $k \in K$, and for each $k \in K - \{0\}$, $\iota(k)$ is chosen from a discrete probability distribution in which the probability that $\iota(k) = i$ is $\rho_i(k)$.

Definition 1.3. Populations, Occupancy, and Population Density Let $\Theta = (V, K, T, \boldsymbol{\rho})$ be a transition model, and let I be a set of n individuals on Θ. The *node occupancy* $N(I, i, k)$ of node i and time k for the population I is defined as

$$N(I, i, k) = |\{\iota \in I \mid \iota(k) = i\}|$$

The *occupancy vector* $\boldsymbol{o}(I, k)$ is the vector such that $o_i(I, k) = N(I, i, k)$, which we note must sum to n. The *population density* $\boldsymbol{p}(k)$ of the set of individuals I at time $k \in K$ is the column vector $\boldsymbol{p}$, where

$$p_i = \frac{1}{n} N(I, ik)$$

The set I is called a *population* if its population density $\boldsymbol{p}(0) = \boldsymbol{\rho}(0)$.

Proposition 1.1. Let $\Theta = (V, K, T, \boldsymbol{\rho})$ be a transition model. For any $t \in K$,

$$\boldsymbol{\rho}(t) = \left(\prod_{k=0}^{t-1} T(t-1-k) \right) \boldsymbol{\rho}(0)$$

Proof: If $t = 0$, the proof is trivial. Otherwise, assume the proposition is true for some $t = t'$. Then if $t = t' + 1$, we have

$$\begin{aligned}
\boldsymbol{\rho}(t'+1) &= T(t')\boldsymbol{\rho}(t') \\
&= T(t') \left(\prod_{k=0}^{t'-1} T(t'-1-k) \right) \boldsymbol{\rho}(0) \\
&= \left(\prod_{k=0}^{t'} T(t'-k) \right) \boldsymbol{\rho}(0)
\end{aligned}$$

Thus the proposition follows by induction.

1.2 Statistics of Populations

All individuals in a population I on a transition model $\Theta = (V, K, T, \boldsymbol{\rho})$ constitute independent discrete random variables, where outcomes V are chosen with respective probabilities $\boldsymbol{\rho}(k)$. Thus, the population's occupancy vector $\boldsymbol{o}(k)$ is a random variable chosen from a multinomial distribution. From this observation, we immediately note several properties of the statistics of a population's occupancy:

Proposition 1.2. Let $\Theta = (V, K, T, \boldsymbol{\rho})$ be a transition model, and let I be a population of n individuals on Θ. The following are true:

1. The probability of obtaining a node occupancy vector $\boldsymbol{o}(k)$ from I is

$$\mathrm{P}(\boldsymbol{o}(k) \mid \Theta) = n! \prod_{i=1}^{|V|} \frac{\rho_i(k)^{o_i}}{o_i!}$$

2. The mean occupancy for each node is given by $\langle \boldsymbol{o}(k) \rangle = n\boldsymbol{\rho}(k)$.

3. The covariance matrix for this occupancy distribution is

$$\boldsymbol{\Sigma}_{ij}(k) = \begin{cases} n\rho_i(k)(1-\rho_i(k)) & i = j \\ n\rho_i(k)\rho_j(k) & i \neq j \end{cases}$$

4. When n is large, we can approximate the $\boldsymbol{o}(k)$ distribution by the normal distribution $\mathrm{P}(\boldsymbol{o}(k)\,|I\,\Theta) \approx \mathcal{N}(\boldsymbol{\rho}(k), \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is the covariance matrix defined above.

Proof: These are all properties of a multinomial distribution.

1.3 Transition Model Fitting

Suppose we have a set of nodes V and times K, as well as occupancy vector observations $\boldsymbol{o}(k)$ for each $k \in K$, and we wish to fit a transition model $\Theta = (V, K, T, \boldsymbol{\rho})$ to explain these occupancy observations. We would be wise to find choose Θ so that the probability of these observations $\boldsymbol{o}(k)$ given Θ is maximized. Since each observation is an independent event, we compute this probability as

$$\mathrm{P}(\boldsymbol{o}(0), \boldsymbol{o}(1), \ldots, \boldsymbol{o}(L)|\Theta) = \prod_{k=0}^{L} \mathrm{P}(\boldsymbol{o}(k)|\Theta)$$

Fortunately, we can simplify this optimization problem by maximizing the log likelihood function

$$\begin{aligned}
\log \mathcal{L} &= \sum_{k=0}^{|K|-1} \log \mathrm{P}(\boldsymbol{o}(k)|\Theta) \\
&= \sum_{k=0}^{|K|-1} \log \left[n! \prod_{i=1}^{|V|} \frac{\rho_i(k)^{o_i(k)}}{o_i(k)!} \right] \\
&= \sum_{k=0}^{|K|-1} \log n! + \sum_{i=1}^{|V|} o_i(k) \log \rho_i(k) - \log o_i(k)! \\
&= C + \sum_{k=0}^{|K|-1} \sum_{i=1}^{|V|} o_i(k) \log \rho_i(k)
\end{aligned}$$

where C is a constant function of $n!$ and $\boldsymbol{o}(k)$. This leads to the following important result:

Proposition 1.3. Let $\boldsymbol{o}(k)$ be a sequence of observation vectors over a set of nodes V and times K. To find a transition model $\Theta = (V, K, T, \boldsymbol{\rho})$ maximizing the log likelihood of these observations, it is sufficient to find Θ that maximizes the quantity

$$\ell = \sum_{k=0}^{|K|-1} \sum_{i=1}^{|V|} o_i(k) \log \rho_i(k)$$

Suppose that the transition matrix T is determined by a set of parameters $\{a_1, a_2, \ldots, a_m\}$. Then we compute

$$\frac{\partial \ell}{\partial a_s} = \sum_{t=0}^{|K|-1} \sum_{n=1}^{|V|} \frac{\partial \ell}{\partial \rho_n(t)} \frac{\partial \rho_n(t)}{\partial a_s}$$

where each population density $\rho_n(t)$ varies with a_s as

$$\frac{\partial \rho_n(t)}{\partial a_s} = \sum_{k=0}^{|K|-1} \sum_{i=1}^{|V|} \sum_{j=1}^{|V|} \frac{\partial \rho_n(t)}{\partial T_{ij}(k)} \frac{\partial T_{ij}(k)}{\partial a_s}$$

and $\rho_n(t)$ varies with the transition matrix entry $T_{ij}(k)$ as

$$\frac{\partial \rho_n(t)}{\partial T_{ij}(k)} = \sum_{m=1}^{|V|} \frac{\partial}{\partial T_{ij}(k)} \left[T_{nm}(t-1) \rho_m(t-1) \right]$$

Note that the population density $\rho_n(t)$ has no dependence on $T_{ij}(k)$ if $k > t-1$. Furthermore, if $k = t-1$, then the above expression reduces to $\rho_j(t-1)$. Noting these simplifications, we can write

$$\frac{\partial \rho_n(t)}{\partial T_{ij}(k)} = \begin{cases} 0 & k+1 > t \\ p_j(t-1) & k+1 = t \\ \sum_{m=1}^{|V|} T_{nm}(t-1) \frac{\partial \rho_m(t-1)}{\partial T_{ij}(k)} & k+1 < t \end{cases}$$

Such a recursive definition for $\partial \rho_n(t)/\partial T_{ij}(k)$ facilitates the computation of a gradient for ℓ with respect to the parameters a_s using dynamic programming. Once such a gradient function is constructed, an optimization strategy like gradient descent can find the parameters to optimize ℓ.