Modeling Markov Chains with Point Processes

Jacob Mortensen, Simon Fraser University
Luke Bornn, Simon Fraser University

Abstract:

When considering movement in space, a useful tool is a Markov model, where the position of the agent at time t+1 depends only on their position at time t. We build on existing theory to show that as the number of spatial locations approaches infinity, a Markov model can be represented by a Poisson point process, a popular type of spatial model that accounts for correlation between nearby locations. Using SportVu player tracking data provided by the National Basketball Association we show how this model can be used to produce distinct maps of player movement for each team in the NBA.

Theoretical Overview:

Consider a first order Markov chain that has values in some domain $\mathcal{D} \subseteq \mathbb{R}$ at times $t = 0, \ldots, T$. One simple way to analyze this data using a Markov model is to partition $\mathcal{D}$ into K states $A_1, \ldots, A_K$, then calculate the transition probabilities $p_{ij} = P(X_t \in A_j | X_{t-1} \in A_i)$ for all $j, i = 1, \ldots, K$ Let the transition probability matrix

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1K} \\ \vdots & \ddots & \vdots \\ p_{K1} & \cdots & p_{KK} \end{bmatrix}$$

where the rows sum to 1 so that the values in row k represent the probability of moving from state A_k to each other state. Given the initial observation X_0, the conditional likelihood for the Markov chain is

$$L(P) = \prod_{t=1}^{T} p_{x_{t-1}, x_t} = \prod_{i=1}^{K} \prod_{j=1}^{K} p_{ij}^{N_{ij}}$$

where, in order to simplify notation, $p_{x_{t-1}, x_t} = p_{ij}$ such that $x_{t-1} \in A_i$ and $x_t \in A_j$ and where

$$N_{ij} = \sum_{t=1}^{T} \mathbb{1}\left[x_{t-1} \in A_i, x_t \in A_j\right] \quad .$$

Note that the conditional likelihood listed above is equivalent to the product of K independent multinomial likelihoods, and so we can estimate the transition probabilities using a multinomial distribution.

Now suppose $\mathbf{y} = (y_1, \ldots, y_K)$ are independent Poisson random variables with means $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$. Then given $n = \sum_{j=1}^{k} y_j$,

$$f(y|n, \boldsymbol{\lambda}) = \prod_{j=1}^{K} \frac{\exp(-\lambda_j)\lambda_j^{y_j}}{y_j!} \frac{n!}{\exp\left(-\sum_{j=1}^{K} \lambda_j\right)\left(\sum_{i=1}^{K} \lambda_i\right)^n}$$

$$= \frac{n!}{y_1! y_2! \ldots y_K!} \prod_{j=1}^{K} \left(\frac{\lambda_j}{\sum_{i=1}^{K} \lambda_j}\right)^{y_j},$$

which is the likelihood for the $\text{Multinomial}(y|n; \alpha_1, \ldots, \alpha_K)$ distribution, where $\alpha_j = \lambda_j / \sum_{i=1}^{K} \lambda_j$. The parallel to our problem is obvious; If we let $N_{i\cdot} = \sum_{j=1}^{K} N_{ij}$ and assume $N_{ij} \sim \text{Poisson}(\lambda_{ij})$ then we can estimate the underlying transition probabilities by letting $p_{ij} = \lambda_{ij} / \sum_{j=1}^{K} \lambda_{ij}$. Next we show that as the number of states in a bounded domain approaches infinity, our Poisson distribution becomes a Poisson process.

In our initial definition we focus on point processes in $\mathbb{R}$ but note that point processes easily generalize to higher dimensions (as demonstrated by our basketball example). A point process is a stochastic process over a domain $\mathcal{D}$ where realizations consist of a finite set of points $\mathcal{S} = \{s_1, \ldots, s_n : s_i \in \mathcal{D}\}$ whose distribution is governed by an intensity function $\Lambda(s)$. A point process is said to be Poisson if, for any subset $\mathcal{B} \subseteq \mathcal{D}$, the random number of points falling in $\mathcal{B}$, $N(\mathcal{B})$, is distributed $\text{Poisson}(\delta)$, where $\delta = \int_{\mathcal{B}} \Lambda(s) ds$. Additionally, if $\mathcal{B}_1$ and $\mathcal{B}_2$ are independent, then $N(\mathcal{B}_1)$ and $N(\mathcal{B}_2)$ are also independent (Moller and Waagepetersen, 2003; Bannerjee et. al., 2014). The likelihood for the Poisson point process is

$$L(\Lambda(s), s \in \mathcal{D}; s_1, \ldots, s_n) = \prod_i \Lambda(s_i) \exp\left(-\Lambda(\mathcal{D})\right),$$

where $\Lambda(\mathcal{D}) = \int_{\mathcal{D}} \Lambda(s) ds$. If we partition $\mathcal{D}$ into a series of discrete intervals $A_1, \ldots, A_K$ then by the Poisson assumption, the likelihood is a product of independent Poisson random variables:

$$\prod_k \exp\left(-\Lambda(A_k)\right) \left(\Lambda(A_k)\right)^{N(A_k)} / N(A_k)!.$$

As the partition grows increasingly fine, $N(A_k) = 1$ or 0 depending on whether or not there is an observation in A_k and in the limit, this likelihood becomes the point process likelihood. Returning to our Markov chain example and recalling that it can be represented using the Poisson distribution, if we think of λ_{ij} as an integrated value from some intensity $\Lambda(\cdot)$, it follows that as the number of states increases we can estimate a transition probability surface using a Poisson point process. We choose to model Poisson point processes using a log Gaussian Cox process (LGCP), i.e. $\Lambda(s) = \exp(Z(s))$ where $Z(s)$ is a Gaussian process. We fit LGCP's using the INLA software package (Lindgren and Rue, 2015) in R (R Core Team, 2016).

One Dimensional Example:

We simulate a first order Markov chain by drawing $X_0 \sim U(0, 1)$ and then drawing $X_t \sim \mathcal{N}_{[0,1]}\left(X_{t-1}, 0.15^2\right)$, where $\mathcal{N}_{[0,1]}$ represents the truncated normal distribution on $[0, 1]$, for $t = 1, \ldots, 1000$. The trace plot for the simulated chain is shown in Figure 1. Because we have both an origin value and destination value, we fit an LGCP in two dimensions. Then we normalize the estimated intensity surface by dividing all points that share an x_{t-1} value by their sum (i.e., dividing each "row" by its sum) to get a transition probability surface. Note that this is only possible because we are plotting a discrete approximation to the continuous surface.

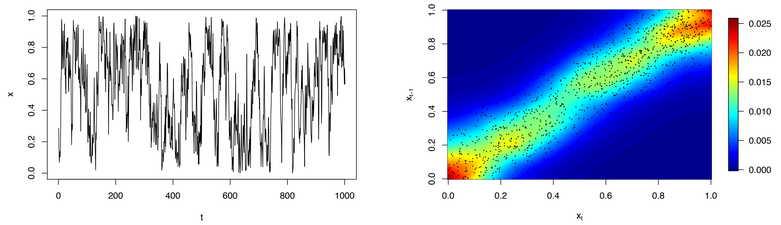

Figure 1. The plot on the left shows the trace plot for the randomly generated Markov chain over time. The right plot shows the estimated transition surface with the (x_t, x_{t-1}) coordinate pairs on top. Note that the y-axis is x_{t-1} in order to be consistent with the TPM.

NBA SportVu Data:

The NBA collects xy-coordinate data for all 10 players on the court and the ball at the rate of 25 measurements per second. If we want to understand how teams move the ball on the court, we can assume that the movement is Markovian and estimate transition surfaces using a 4 dimensional LGCP. By conditioning on a specific point on the court and normalizing the surface we get a transition probability surface from that point to all other points on the court. Because we calculate these surfaces for each team, we can gain insight into how teams use the court space differently. Below we see transition probability surfaces for the Cleveland Cavaliers (left) and the Golden State Warriors (right) for the 2016 regular season, conditioned on the point marked by the purple x. While at first glance these exhibit great similarity, by subtracting the Warriors' surface from the Cavaliers' (bottom) we see that, from this location, the Cavs are more likely to stay on the perimeter and within the immediate vicinity of the x than the Warriors, while the Warriors are more likely to move the ball directly forward.

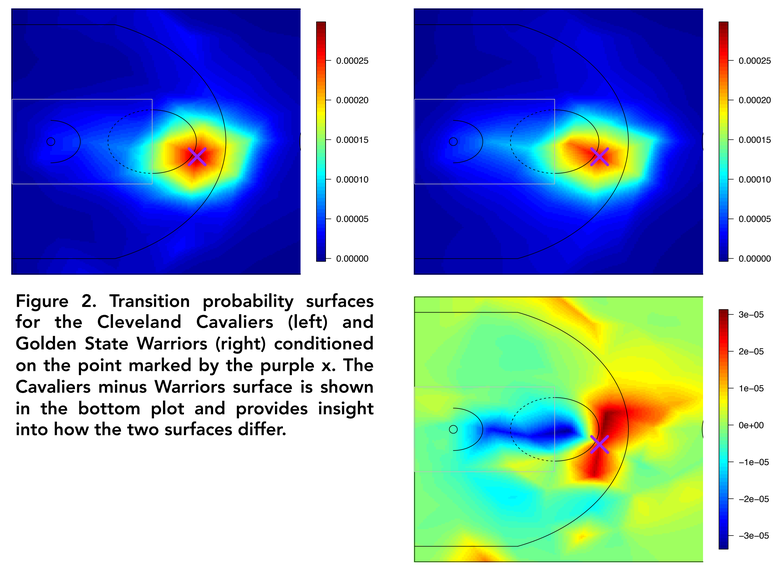

Figure 2. Transition probability surfaces for the Cleveland Cavaliers (left) and Golden State Warriors (right) conditioned on the point marked by the purple x. The Cavaliers minus Warriors surface is shown in the bottom plot and provides insight into how the two surfaces differ.