STAT 870 Lecture 12

Hitting Times

Start irreducible recurrent chain X_n in state i. Let T_j be first $n > 0$ such that $X_n = j$. Define

$$m_{ij} = \mathrm{E}(T_j | X_0 = i)$$

First step analysis:

$$\begin{aligned} m_{ij} &= 1 \cdot P(X_1 = j | X_0 = i) \\ &\quad + \sum_{k \neq j} (1 + \mathrm{E}(T_j | X_0 = k)) P_{ik} \\ &= \sum_j P_{ij} + \sum_{k \neq j} P_{ik} m_{kj} \\ &= 1 + \sum_{k \neq j} P_{ik} m_{kj} \end{aligned}$$

Example

$$\mathbf{P} = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

The equations are

$$\begin{aligned} m_{11} &= 1 + \frac{2}{5} m_{21} \\ m_{12} &= 1 + \frac{3}{5} m_{12} \\ m_{21} &= 1 + \frac{4}{5} m_{21} \\ m_{22} &= 1 + \frac{1}{5} m_{12} \end{aligned}$$

The second and third equations give immediately

$$\begin{aligned} m_{12} &= \frac{5}{2} \\ m_{21} &= 5 \end{aligned}$$

Then plug in to the others to get

$$m_{11} = 3$$
$$m_{22} = \frac{3}{2}$$

Notice stationary initial distribution is

$$\left(\frac{1}{m_{11}}, \frac{1}{m_{22}}\right)$$

Consider fraction of time spent in state j:

$$\frac{1(X_0 = j) + \cdots + 1(X_n = j)}{n+1}$$

Imagine chain starts in chain i; take expected value.

$$\frac{\sum_{r=1}^{n} \mathbf{P}_{ij}^{r} + 1(i = j)}{n+1}$$

If rows of $\mathbf{P}$ converge to π then fraction converges to π_j; i.e. limiting fraction of time in state j is π_j.

Heuristic: start chain in i. Expect to return to i every m_{ii} time units. So are in state i about once every m_{ii} time units; i.e. limiting fraction of time in state i is $1/m_{ii}$.

Conclusion: for an irreducible recurrent finite state space Markov chain

$$\pi_i = \frac{1}{m_{ii}}.$$

Real proof: Renewal theorem or variant.

Idea: $S_1 < S_2 < \ldots$ are times of visits to i. Segment i:

$$X_{S_{i-1}+1}, \ldots, X_{S_i}.$$

Segments are iid by Strong Markov.

Number of visits to i by time S_k is exactly k.

Total elapsed time is $S_k = T_1 + \cdots + T_k$ where T_i are iid.

Fraction of time in state i by time S_k is

$$\frac{k}{S_k} \to \frac{1}{m_{ii}}$$

by SLLN. So if fraction converges to π_i must have

$$\pi_i = \frac{1}{m_{ii}}.$$

Summary of Theoretical Results:

For an irreducible aperiodic positive recurrent Markov Chain:

1. $\mathbf{P}^n$ converges to a stochastic matrix $\mathbf{P}^\infty$.
2. Each row of $\mathbf{P}^\infty$ is π the unique stationary initial distribution.
3. The stationary initial distribution is given by

   $$\pi_i = 1/m_i$$

   where m_i is the mean return time to state i from state i.

If the state space is finite an irreducible chain is positive recurrent.

Ergodic Theorem

Notice slight of hand: I showed

$$\frac{\mathrm{E}\left\{\sum_{i=0}^n 1(X_i = k)\right\}}{n} \to \pi_i$$

but claimed

$$\frac{\sum_{i=0}^n 1(X_i = k)}{n} \to \pi_i$$

almost surely which is also true. This is a step in the proof of the ergodic theorem. For an irreducible positive recurrent Markov chain and any f on S such that $\mathrm{E}^\pi(f(X_0)) < \infty$:

$$\frac{\sum_0^n f(X_i)}{n} \to \sum \pi_j f(j)$$

almost surely. The limit works in other senses, too. You also get

$$\frac{\sum_0^n f(X_i, \ldots, X_{i+k})}{n} \to \mathrm{E}^\pi \left\{f(X_0, \ldots, X_k)\right\}$$

E.g. fraction of transitions from i to j goes to

$$\pi_i \mathbf{P}^{ij}$$

For an irreducible positive recurrent chain of period d:

1. $\mathbf{P}^d$ has d communicating classes each of which forms an irreducible aperiodic positive recurrent chain.

2. $(\mathbf{P}^{n+1} + \cdots + \mathbf{P}^{n+d})/d$ has a limit $\mathbf{P}^\infty$.

3. Each row of $\mathbf{P}^\infty$ is π the unique stationary initial distribution.

4. Stationary initial distribution places probability $1/d$ on each of the communicating classes in 1.

For an irreducible null recurrent chain:

1. $\mathbf{P}^n$ converges to 0 (pointwise).

2. there is no stationary initial distribution.

For an irreducible transient chain:

1. $\mathbf{P}^n$ converges to 0 (pointwise).

2. there is no stationary initial distribution.

For a chain with more than 1 communicating class:

1. If $\mathcal{C}$ is a recurrent class the submatrix $\mathbf{P}_\mathcal{C}$ of $\mathbf{P}$ made by picking out rows i and columns j for which $i, j \in \mathcal{C}$ is a stochastic matrix. The corresponding entries in $\mathbf{P}^n$ are just $(\mathbf{P}_\mathcal{C})^n$ so you can apply the conclusions above.

2. For any transient or null recurrent class the corresponding columns in $\mathbf{P}^n$ converge to 0.

3. If there are multiple positive recurrent communicating classes then the stationary initial distribution is not unique.

Poisson Processes

Particles arriving over time at a particle detector. Several ways to describe most common model.

Approach 1: numbers of particles arriving in an interval has Poisson distribution, mean proportional to length of interval, numbers in several non-overlapping intervals independent.

For $s < t$, denote number of arrivals in $(s, t]$ by $N(s, t)$. Jargon: $N(A) =$ number of points in A is a **counting process**. Model is

1. $N(s,t)$ has a Poisson($\lambda(t-s)$) distribution.

2. For $0 \le s_1 < t_1 \le s_2 < t_2 \cdots \le s_k < t_k$ the variables $N(s_i,t_i); i = 1,\ldots,k$ are independent.

Approach 2:
Let $0 < S_1 < S_2 < \cdots$ be the times at which the particles arrive.
Let $T_i = S_i - S_{i-1}$ with $S_0 = 0$ by convention. T_i are called **interarrival** times.
Then $T_1, T_2, \ldots$ are independent Exponential random variables with mean $1/\lambda$.
Note $P(T_i > x) = e^{-\lambda x}$ is called **survival function** of T_i.
Approaches are equivalent. Both are deductions of a model based on **local** behaviour of process.
Approach 3: Assume:

1. given all the points in $[0,t]$ the probability of 1 point in the interval $(t,t+h]$ is of the form
$$\lambda h + o(h)$$

2. given all the points in $[0,t]$ the probability of 2 or more points in interval $(t,t+h]$ is of the form
$$o(h)$$

Notation: given functions f and g we write
$$f(h) = g(h) + o(h)$$
provided
$$\lim_{h\to 0} \frac{f(h)-g(h)}{h} = 0$$
[Aside: if there is a constant M such that
$$\limsup_{h\to 0} \left|\frac{f(h)-g(h)}{h}\right| \le M$$
we say
$$f(h) = g(h) + O(h)$$
Notation due to Landau. Another form is
$$f(h) = g(h) + O(h)$$

means there is $\delta > 0$ and M s.t. for all $|h| < \delta$

$$|f(h) - g(h)| \le Mh$$

Idea: $o(h)$ is tiny compared to h while $O(h)$ is (very) roughly the same size as h.]

Generalizations:

1. First (Poisson) model generalizes to $N(s,t]$ having a Poisson distribution with parameter $\Lambda(t) - \Lambda(s)$ for some non-decreasing non-negative function Λ (called **cumulative intensity**). Result called **inhomogeneous** Poisson process.

2. Exponential interarrival model generalizes to independent non-exponential interarrival times. Result is **renewal process** or **semi-Markov** process.

3. Infinitesimal probability model generalizes to other infinitesimal jump rates. Model specifies **infinitesimal generator**. Yields other **continuous time Markov Chains**.