Lecture 11: Markov Chains II

We first finished off our study of point to point limits ($p_{ij}(n)$ as $n->\infty$) and then moved on to talk about limit distributions.

Point to Point

We began with the delayed

Proof of 6.3.2(c): We actually prove that if i is positive persistent so is j which is equivilant by 6.3.2(b). So we assume μ_i is finite and prove μ_j is.

We let μ_{ij} be $\sum_n nf_{ij}(n)$ and define μ_{ji} similarily. Since, $\mu_{ij}+\mu_{ji}$ measures the expected time to visit i and THEN return to j it is at least μ_j. We will prove μ_j is finite by proving that μ_{ji} and μ_{ij} are.

To this end we let $p_{j<i}$ be probability of the event that we visit j before returning to i if we start at i. Since $i<->j$ this exceeds 0. If $p_{j<i}$ is 1, then $\mu_i=\mu_{ij}+\mu_{ji}$ and we are done. So we can assume $p_{j<i}$ lies strictly between 0 and 1.

Now, conditioned on visiting j before returning to i, the expected time to return to i is at least the time it takes us once we have arrived at j for the first time which is μ_{ji}. So, the contribution to μ_i from walks in this event exceeds $p_{j<i}\mu_{ji}$. Since μ_i is finite and $p_{j<i}>0$, we see that μ_{ji} is finite.

Now, we let μ_i^1 be the expected time to return to i given we do not visit j on the way, and μ_i^2 be the expected time to return to i given we go through j. Then, $\mu_i=p_{j<i}\mu_i^2+(1-p_{j<i})\mu_i^1$. Since μ_i is finite and both probabilities in the equation are non zero, we see that μ_i^1 and μ_i^2 are both finite.

Now, by homogeneity, $\mu_{ij}\le p_{j<i}\mu_i^2+(1-p_{j<i})(\mu_i^1+\mu_{ij})$. Solving, we see that $\mu_{ij}\le\mu_i^2+\frac{1-p_{j<i}}{p_{j<i}}\mu_i^1$. It follows that μ_{ij} is finite.

QED.

Next we proved that, assuming 6.2(9) which we have yet to prove, if j is transient or null persitent then for every i, $p_{ij}(n)$ goes to zero as n goes to infinity.

Proof: Fix an $\epsilon>0$. Since the probability we visit j when starting at i is at most one, we can choose N so that $\sum_{n>N}f_{ij}(n)$ is at most $\frac{\epsilon}{2}$. By 6.2(9), we can choose N' so that for $n>N'$, $p_{jj}(n)\le\frac{\epsilon}{2}$.

Now, for $n\ge N+N'$, we have:

$$p_{ij}(n)=\sum_{k\le n}f_{ij}(k)p_{jj}(n-k)=\sum_{k\le N}f_{ij}(k)p_{jj}(n-k)+\sum_{N<k\le n}f_{ij}(k)p_{jj}(n-k)$$

$$\le\sum_{k\le N}f_{ij}(k)\frac{\epsilon}{2}+\sum_{N<k}f_{ij}(k)\le\frac{\epsilon}{2}+\frac{\epsilon}{2}.$$

QED.

Limit Distributions

We started to talk about limit distributions. We defined limit distributions in terms of Total Variational Distance (defined on page 44 of the text). We noted that a limit distribution must be stationary (i.e. $xP = x$). Using 6.2(9) we proved that a stationary distribution is 0 on transient or null-persistent states. We noted that stationary distributions on a Markov chain were convex combinations of the stationary distributions on the subchains given by the equivalence classses of postive persistent states. We noted that to have a limit distribution, a chain must have at most one equivalence class consisting of positive persistent states, and these states must be aperiodic. Since this would also be a limit distribution for the submarkov chain given by the equivalence class. It seems natural to restrict our attention to this subchain.

Thus limit distributions are really of interest only for closed, irreducible, aperiodic markov chains whose states are positive persistent.

We proved (actually we will prove next class) that the limit distribution for such a chain is given by $x(i) = \frac{1}{\mu_i}$. The first step is to show that this is the only candidate for a stationary distribution.

Suppose x is a stationary distribution and fix a state j. If we start our chain in $x_0 = x$ then $x_n(j) = x(j)$ for every n. We want to prove this is $\frac{1}{\mu_j}$. To this end, choose N s.t. $\sum_{i \leq N} x(i) \geq 1 - \frac{\epsilon}{2}$. Since $p_{ij}(n)$ goes to $\frac{i}{\mu_j}$ for every n, we can choose N' such that for $i \leq N$ and $n \geq N'$ $|p_{ij}(n) - \frac{1}{\mu_j}| \leq \frac{\epsilon}{2}$.

Now, for any $n > N + N'$, $x_n(j) = \sum_i x(i)p_{ij}(n)$. That is:

$$x(j) - \frac{1}{\mu_j} = \sum_{i \leq N} x(i)(p_{ij}(n) - \frac{1}{\mu_j}) + \sum_{i > N} x(i)(p_{ij}(n) - \frac{1}{\mu_j}).$$

Since the absolute value of a sum is at most the sum of the asbsolute values, this yields:

$$|x(j) - \frac{1}{\mu_j}| = \sum_{i \leq N} x(i)\frac{\epsilon}{2} + \sum_{i > N} x(i)|p_{ij}(n) - \frac{1}{\mu_j})|.$$

Which since x is a probability distribution, yields by our choice of N':

$$|x(j) - \frac{1}{\mu_j}| \leq \frac{\epsilon}{2} + \sum_{i > N} x(i) < \epsilon.$$

Since this holds for all positive ϵ, $x(j)$ is indeed $\mu(j)$.

Our next step is to show that the inverses of the mean times between visits do indeed yield a stationary distribution. We follow the argument on page 228 in the text to show that there is a stationary distribution x. By our above remarks, (j) must be $\frac{1}{\mu_j}$ for each j. It remains to show that for any initial state x_o we have that $TVD(x_n, x)$ goes to zero.

Proof: Fix $\epsilon > 0$. Choose N so that $\sum_{i>N} x(i) < \frac{\epsilon}{6}$ and $\sum_{i>N} x_o(i) < \frac{\epsilon}{6}$. Since $p_{ij}(n)$ goes to $x(j)$ for each j, we can choose N' so that for all $i, j \leq N$ and $n \geq N'$ we have $|p_{ij}(n) - x(j)| \leq \frac{\epsilon}{6N}$.

For $n \geq N'$ and $j \leq N$, we have:

$$x_o P^n(j) = \sum_i x_o(i) p_{ij}(n) \geq \sum_{i \leq N} x_o(i) p_{ij}(n)$$

$$\geq \sum_{i \leq N} x_0(i)(x(j) - \frac{\epsilon}{6N}) \geq (1 - \frac{\epsilon}{6})(x(j) - \frac{\epsilon}{6N}) \geq x(j) - \frac{\epsilon}{6} x(j) - \frac{\epsilon}{6N}$$

Thus

$$TVD(x_o P^n, x) = 2 \sum_j max(0, x(j) - x_o P^n(j))$$

$$\leq 2 \sum_{j<N} (\frac{\epsilon x(j)}{6} + \frac{\epsilon}{6N}) \quad + 2 \sum_{j>N} x(j) \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3}.$$

QED.

The proof of 6.2(9)

We now give the proof that for a null persistent i, $p_{ii}(n)$ goes to zero as n goes to infinity.

The proof had two parts. We proved

(A) for every m and $\epsilon > 0$ there was an N such that for $n \geq N$ and $m \leq k$, $|p_{ii}(n) - p_{ii}(n-k)| \leq \frac{\epsilon}{4}$.

and

(B) for every ϵ there is a d such that for every m there is an N' so that for $n \geq N'$ we have: $\sum_{k=n}^{n-m+1} p_{ii}(k) \leq d + \frac{\epsilon m}{4}$.

The rest was easy. For $\epsilon > 0$ we found d as in (B) and we set $m = \frac{4d}{\epsilon}$. We chose N and N' for this m and ϵ. Then for $n \geq N$, we have by our choice of m and (B) that $\sum_{k=n}^{n-m+1} p_{ii}(k) \leq \frac{\epsilon m}{2}$. Thus, there is an l between n and $n+m-1$ with $p_{ii}(l) \leq \frac{\epsilon}{2}$. By (A), $p_{ii}(n) \leq \epsilon$, and we are done.

To prove (A) and (B) we considered the renewal process where the interarrival time was the random time X it takes the chain to return to i.

To prove (A) we just follow the argument in the book which works to couple two renewal processes with differring starting points regardless of whether the mean step size is finite.

To prove (B), we defined a truncated step distribution X^b for each b such that for $i < b, Prob(X^b = i) = Prob(X = i)$ and for $i > b, Prob(X^b = i) = 0$ and $Prob(X^b = b) = Prob(X \geq b)$. We noted that as b goes to infinity $E(X^b)$ goes to $E(X)$ which is infinite. So, we could and did choose d with $E(X^d) > \frac{8}{\epsilon}$. Clearly, $E(X^d)$ is finite as it is at most d. Thus, we can apply our renewal process result to obtain that as n goes to infinity $p^d_{ii}(n)$ goes to $\frac{1}{E(X^d)}$. So for some N',we have that for $n \geq N'$, $p^d_{ii}(n) \leq \frac{\epsilon}{4}$ and

$$\sum_{k=n}^{n+m-1} p^d_{ii}(k) \leq \frac{\epsilon m}{4}.$$

To complete the proof of B, we need only show that for any a and b we have:

$$\sum_{k=a}^{b-d} p_{ii}(k) \leq \sum_{k=a}^{b} p^d_{ii}(k).$$

To this end we couple the renewal in this range. We can assume there is a renewal of the untruncated process between a and $b - d$ as otherwise there is nothing to prove. This implies that $b \geq a + d$ and hence the first renewal of the truncated process at time a or later (which occurs before time $a + d$) occurs before b.

We can couple the interarrival times starting at this pair of renewals so that either they are the same and less than d, or the truncated process interarrival time is d and the untruncated process interaarival time is at least as big. It follows that in order to move $b+d$ time units, we take more steps in the truncated process than in the untruncated one. QED.