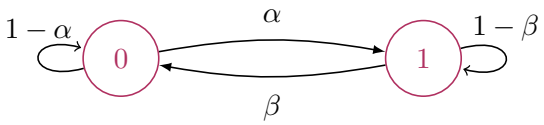
2 EXISTENCE AND CONVERGENCE TO STATIONARITY

Recall from earlier Example were we considered a Markov chain X_n on $S = \{0, 1\}$ with $P(X_0 = 0) = \mu_0 \in (0, 1), P(X_0 = 1) = \mu_1 \in (0, 1)$, and whose dynamics is given by directed graph with S as the vertex set.



with $0 < \alpha, \beta < 1$. We saw that

$$\sup_{i\in\{0,1\}} \mid \mathbb{P}(X_n = i) - \pi(\{i\}) \mid \leq C(1 - \alpha - \beta)^n,$$

with the stationary distribution π given by $\pi(\{0\}) = \frac{\beta}{\alpha+\beta}$ and $\pi(\{1\}) = \frac{\alpha}{\alpha+\beta}$. While not every Markov chain may have a stationary distribution, it is nevertheless true that every irreducible Markov chain on a finite state space S has a unique stationary distribution. Our treatment is borrowed from [FW, Lemma 3.1 Page 158] and [LPW, Lemma 1.16, Theorem 4.9].

THEOREM 2.1. *Let X_n be an irreducible Markov chain on a finite state space S. Then X_n has a unique stationary distribution.*

Proof : We shall first show existence of a stationary distribution. For $i \in S$, a directed spanning tree on S with root at i is by definition a graph g which satisfies the following conditions:

(a) g has verices indexed by S,

(b) the edges of g are directed in such a way that for any $j \in S$, there is a unique directed path from j to i,

(c) $i \in S$ does not have outgoing edge and $j \in S \setminus \{i\}$ has one outgoing edge.

Let $G(i)$ be the set of all directed spanning trees on S with root at i. For every $i \in S$, define

$$\pi_0(i) = \sum_{g\in G(i)} \prod_{(j\to k)\in g} p_{jk}, \tag{2.1}$$

where $j \to k$ denotes any directed edge in the tree g. As X_n is irreducible each $\pi_0(i) > 0$. We shall first verify that

$$\pi_0(i) = \sum_{j\in S} \pi_0(j)p_{ji}, \forall i \in S,$$

which is equivalent to showing that

$$\pi_0(i)\sum_{j\neq i} p_{ij} = \sum_{j\neq i}\pi_0(j)p_{ji} \tag{2.2}$$

or that

$$\Big(\sum_{g\in G(i)}\prod_{(j\to k)\in g} p_{jk}\Big)\sum_{j\neq i} p_{ij} = \sum_{j\neq i}\Big(\sum_{g\in G(j)}\prod_{(k\to j)\in g} p_{kj}\Big)p_{ji} \tag{2.3}$$

The equality follows from the following observation, which can be easily verified. Both sides of (2.3) are equal to $\sum_{g\in L}\prod_{(j\to k)\in g} p_{jk}$, where L is the set of all graphs satisfying:

(a) every point $k \in S$ has only one directed edge pointing way from k to l, $(l \neq k, l \in S)$

(b) the graph has exactly one closed cycle and $i \in S$ belongs to that cycle.

So the normalised version of π_0 - defined by $\pi(i) = \Big(\sum_{j\in S}\pi_0(j)\Big)^{-1}\pi_0(i)$ - will be a stationary distribution for X_n.

Now we shall prove uniqueness. As S is finite, let $\pi = [\pi_1(1) - \pi_2(1), \pi_1(2) - \pi_2(2), \ldots]$ be thought of as a row vector. As π is stationary we have

$$\pi\mathbf{P} = \pi \qquad \text{equivalently} \qquad \pi(\mathbf{P} - I) = 0 \qquad \text{equivalently} \qquad (\mathbf{P}^T - I)\pi^T = 0$$

We will show that the null space of $P - I$ has dimension 1 which will imply that the null space of $\mathbf{P}^T - I$ is also one dimension which implies that there is only one probability vector π that satisfies $(\mathbf{P}^T - I)\pi^T = 0$.

Let $h : S \to \mathbb{R}$ be a solution to $\mathbf{P}h = h$, with h being thought of as a column vector. Set $M = \max\{h(i) : i \in S\}$ and $h(i_0) = M$ for some $i_0 \in S$. Suppose $p_{i_o j} > 0$ and $h(j) < M$ then

$$h(i_o) = h(j)p_{ji_0} + \sum_{k\in S, k\neq j} h(k)p_{kj_0} < M(p_{i_o j} + \sum_{k\in S, k\neq j} p_{ki_0}) = M.$$

This is a contradiction. Therefore $h(j) = M$. By irreducibility it follows that $h(k) = M$ for any $k \in M$. Hence the

$$\text{null space of } P - I = \left\{ h : h = \alpha \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \alpha \in \mathbb{R} \right\}$$

and has dimension 1. ■

We will now establish a rate of convergence to stationary distribution for an irreducible aperiodic Markov chain on a finite state space.

THEOREM 2.2. *Let X_n be a Markov chain on S with transition matrix $\mathbf{P}$ and initial distribution μ. Let X_n be irreducible and aperiodic Markov chain with $\mid S \mid < \infty$. Let π be its unique stationary distribution. Then there exist $C > 0$ and $0 < \theta < 1$ such that for all $n \geq 1$,*

$$\sup_{i,j \in S} \mid p_{ij}^n - \pi(\{j\}) \mid \leq C \cdot \theta^n \tag{2.4}$$

Proof : Now without loss of generality let $S = \{1, 2, \ldots, N\}$ for some $N \geq 1$. We shall first show that there exists $r \geq 1$ such

$$p_{ij}^r > 0 \qquad \text{for all } i, j \in S. \tag{2.5}$$

Since the chain is aperiodic, for each $i \in S$ there is $n_i \geq 1$ such that $p_{ii}^n > 0$ for all $n \geq n_i$. For each $j \in S$ and $j \neq i$ by irreducibility there is a n_j such that $p_{ij}^{n_j} > 0$. Let $r = \max\{n_i : i \in S\} + \max\{n_j : j \in S\}$. Then for any $i, j \in S$,

$$p_{ij}^r \geq p_{ii}^{r-n_j} p_{ij}^{n_j} > 0,$$

which establishes (2.5). Let $0 < \delta < 1$ be such that

$$p_{ij}^r \geq \delta \pi(\{i\}) > 0,$$

for all $i, j \in S$. Now without loss of generality let $S = \{1, 2, \ldots, N\}$ for some $N \geq 1$. Define the matrix $\Pi_{N \times N}$ by

$$\Pi = \begin{pmatrix} \pi(\{1\}) & \pi(\{2\}) & \ldots & \pi(\{N\}) \\ \pi(\{1\}) & \pi(\{2\}) & \ldots & \pi(\{N\}) \\ \pi(\{1\}) & \pi(\{2\}) & \ldots & \pi(\{N\}) \\ \vdots & \vdots & \vdots & \vdots \\ \pi(\{1\}) & \pi(\{2\}) & \ldots & \pi(\{N\}) \end{pmatrix}.$$

Let $\theta := 1 - \delta$ and define the matrix $Q_{N \times N}$ by

$$Q := \frac{1}{\theta}(P^r + (1 - \theta)\Pi).$$

By construction $Q = (q_{ij})_{1 \leq i,j \leq N}$ is such that $0 \leq q_{ij} \leq 1$ and $\sum_{j=1}^N q_{ij} = 1$ for all $i \in S$. Rearranging the above, using the fact that $Q^k \Pi = \Pi$ and$\Pi P^r = \Pi$ along with induction it is easy to see that for all $k \geq 1$,

$$P^{rk} = \theta^k Q^k + (1 - \theta^k)\Pi.$$

Let $n \geq 1$. Then there is $k \geq 1$ and $0 \leq l \leq r-1$ such that $n = rk + l$. We multiply the above by P^l to get

$$\begin{aligned} P^{rk+l} &= \theta^k Q^k P^l + (1-\theta^k)\Pi P^l \\ &= \theta^k Q^k P^l + (1-\theta^k)\Pi, \end{aligned}$$

as π is as stationary distribution. Therefore, by rearranging we have

$$P^{rk+l} = \theta^k (Q^k P^l - \Pi).$$

If we consider each in the matrix equality above and take absolute values we have that for each $i \in S$ we have

$$p_{ij}^{rk+j} - \pi(\{j\}) = \theta^k((Q^k P^l)_{ij} - \pi(\{j\})), \qquad \text{for all } j \in S.$$

Consequently, we have

$$\sup_{i,j \in S} \mid p_{ij}^n - \pi(\{j\}) \mid = \theta^k \mid (Q^k P^l)_{ij} - \pi(\{j\}) \mid \leq C\theta^n,$$

for some $C > 0$. ■

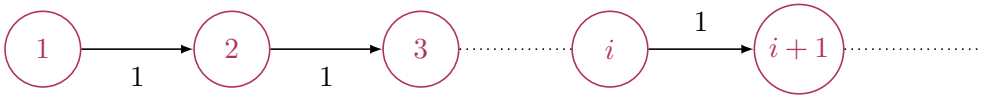
Now observe that, for any $j \in S$ and $n \geq$,

$$\begin{aligned} \mid \mathbb{P}(X_n = j) - \pi(\{j\}) \mid &= \mid \sum_{i \in S} p_{ij}^n \mathbb{P}(X_0 = i) - \pi(\{j\}) \mid \\ &= \sum_{i \in S} \mid p_{ij}^n - \pi(\{j\}) \mid \mathbb{P}(X_0 = i) \\ &\leq \mid S \mid \sup_{i,j \in S} \mid p_{ij}^n - \pi(\{j\}) \mid \end{aligned}$$

Therefore using Theorem 2.2 we have for all $n \geq 1$ there is a $C > 0$ and $0 < \theta < 1$ such that

$$\sup_{j \in S} \mid \mathbb{P}(X_n = j) - \pi(\{j\}) \mid \leq \mid S \mid \cdot C \cdot \theta^n$$

Example 2.3. Let X_n be a Markov chain on state space $\mathbb{N}$ with transition probabilities $p_{ij} = 1$ if $j = i + 1$ (and $p_{ij} = 0$ otherwise). In other words, when $X_n = j$ then $X_{n+1} = j + 1$ with probability one. This chain may be represented by the graph



If π were a stationary distribution for X_n then it would have to be that $\pi(\{i+1\}) = \pi(\{i\})$ for all $i \geq 1$. But since $\sum_{k=1}^{\infty} \pi(\{k\}) = 1$ this is not possible. Therefore this Markov chain has no stationary distribution. ■

We now state some basic properties on stationary distributions and limiting distributions.

PROPOSITION 2.4. *Let X_n be a Markov Chain with transition matrix P on state space S.*

(a) Let $n \in \mathbb{N}$. Suppose π is a stationary distribution of X_n then $X_0 \stackrel{d}{=} \pi$ implies $X_n \stackrel{d}{=} \pi$.

(b) Suppose S is a fininte state space. Then any limiting distribution must be stationary. That is, suppose for some $i \in S$,

$$\lim_{n\to\infty} p_{ij}^{[n]} = \pi(j) \text{ for all } j \in S,$$

then π is a stationary distribution for X_n.

(c) Suppose S is a finite state space and P is doubly stochastic, (i.e $\sum_{i\in S} p_{ij} = 1$ for all $j \in S$) then the uniform distribution on S is stationary. That is, if $\pi(i) = \frac{1}{|S|}$ for all $i \in S$, then π is a stationary distribution for X_n.

Proof :

(a) As π is stationary, this is immediate from (**??**) after an inductive argument.

(b) First let us verify under the hypothesis π is a Probability on S. Clearly, $0 \leq \pi(j) \leq 1$ for all $j \in S$.

$$\sum_{j\in S} \pi(j) = \sum_{j\in S} \lim_{n\to\infty} p_{ij}^{[n]} = \lim_{n\to\infty} \sum_{j\in S} p_{ij}^{[n]} = 1$$

where the limit and sum may be freely interchanged because S is finite. To verify that π is stationary we first observe that for any $n \geq 1$, by (**??**) with $r = n-1$ and $s = 1$ we have

$$p_{ij}^{[n]} = \sum_{k\in S} p_{ik}^{[n-1]} p_{kj}.$$

Letting $n \to \infty$ we have, for all $j \in S$,

$$\pi(j) = \sum_{k\in S} \pi(k) p_{kj}.$$

Hence π is a stationary distribution.

(c) Let π be a probability on S. If $|S| = N$ and $\pi(i) = \frac{1}{N}$ for all $i \in S$, then, for each $j \in S$, observe that

$$\sum_{i \in S} \pi(i) p_{ij} = \frac{1}{N} \sum_{i \in S} p_{ij} = \frac{1}{N} = \pi(j).$$

Also $\sum_{i \in S} \pi(i) = 1$. So π is a stationary distribution of X_n. ■

We shall conclude this discussion with a statement on convergence to stationary for irreducible and aperiodic Markov Chains.

THEOREM 2.5. *Let X_n be a Markov chain on S with transition matrix $\mathbf{P}$ and initial distribution μ. Assume that X_n is aperiodic and irreducible. If X_n has a stationary disrtibution π and if $P(X_0 = i) > 0$, then*

$$\lim_{n \to \infty} p_{ij}^n = \pi(j), \tag{2.6}$$

for all $j \in S$.

REMARK 2.6.

1. Generalisations: If the chain in Theorem 2.5 was periodic with period $d > 1$, it can be shown that there exists an integer r with $0 \leq r < d$ such that

   $$\lim_{m \to \infty} p_{ij}^{md+r} = d\pi(j), \tag{2.7}$$

   for all $j \in S$.

2. Let X_n be a Markov chain on S with transition matrix $\mathbf{P}$ and initial distribution μ. Supppose X_n is a irreducible. It can have atmost one stationary distribution π.

   We have already seen it is true when $\mid S \mid < \infty$. Otherwise this can be reasoned as follows. Assume that X_n has two stationary distributions π_1, π_2. Consider a Markov chain $\widetilde{X}_n$ on state space S with transition matrix $\widetilde{\mathbf{P}} = \frac{I+\mathbf{P}}{2}$ where I is the identity matrix. Note that $\widetilde{X}_n$ is an aperiodic and irreducible Markov chain with stationary distributions π_1 and π_2 (verify this). Now Theorem 2.5 applied to the chain $\widetilde{X}_n$ implies that $\pi_1 = \pi_2$.

3. (Ergodic Theorem) Let X_n be a Markov chain on S with transition matrix $\mathbf{P}$ and initial distribution μ. Assume that X_n is irreducible and aperiodic. For $i, j \in S$, we have

   $$\frac{\sum_{k=1}^n p_{ij}^k}{n} \to \pi(j) \quad a.e. \tag{2.8}$$

   This is known as the Ergodic Theorem for Markov Chains.

BIBLIOGRAPHY

[AS09] Siva, Athreya, Sunder, V.S. *Measure and Probability* CRC Press (Outside India) 2009 ISBN 14 3980 126 6.

[CasBer90] Casella, George; Berger, Roger L.*Statistical inference.* The Wadsworth & Brooks/Cole Statistics/Probability Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1990. xviii+650 pp. ISBN: 0-534-11958-1

[Fel68] W. Feller, *An introduction to probability theory and its applications* Vol. I. Third edition John Wiley & Sons, Inc., New York-London-Sydney 1968.

[Fel71] W. Feller, *An introduction to probability theory and its applications* Vol. II. Second edition John Wiley & Sons, Inc., New York-London-Sydney 1971.

[FPP98] T.Freedman, R.Pisani, R.Purves, *Statistics* Third edition W.W.Norton and Company, Inc., New York-London, 1998.

[FG97] B. Fristedt and L. Gray, *A Modern Approach to Probability Thoery* Birkhauser, Boston 1997.

[FW] M.i. Friedlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Springer-verlag, New York, 1998.

[Ghah00] S. Ghahramani *Fundamentals of probability.* Second edition Prentice Hall, New Jersey, 2000.

[HMC05] R.Hogg, J. McKean, A. Craig, *Introduction to Mathematical Statistics*, Sixth Edition, Pearson Education, Inc., New Jersey, 2005.

[HPS72] P.G. Hoel, S.C. Port, and C.J. Stone, *Introduction to Probability Theory*, Houghton Mifflin Company, 1972.

[Keane] M. Keane, *The Essence of the Law of Large Numbers*, , Pages 125–129, Algorithms, Fractals, and Dynamics, Springer USA, 1995.

[Nor97] J.R. Norris, *Markov Chains,* Cambridge University Press, 1997.

[Pit89] J. Pitman, *Probability,* Springer-verlag, New York, 1989.

[Ram97] Ramasubramanian, *The Normal Distribution 1. From Binomial to Normal* Resonance, Vol 2, No.6, (June 1997), pp 15-14.

[Rao73] C.R.Rao, *Linear Statistical Inference and its Application* , Second Edition, John Wiley, New York, 1973.

[Ross84] S. Ross *A first course in probability*, Second edition Macmillan Publishing Company, New York, 1984.

[Stig84] Stephen M. Stigler, *Kruskal's Proof of the Joint Distribution of* $\overline{X}_n$ *and* s_n^2, The American Statistician, Vol 38, No. 2, (May 1984), pp 134-135.

[LPW] Levin DA, Peres Y. and Wilmer EL, *Markov chains and mixing times.* American Mathematical Soc., Providence. 2009;107.