Thesis Background

Ali Taqi

- .

Notational Conventions

- State spaces will usually be denoted by $S_M = \{1, 2, \ldots, M\}$. One could say that $S_M \subseteq \mathbb{N}$, taking the first M elements. One may have seen this notated before as $[n] = \{k \in \mathbb{N} : k \leq n\}$, but we will stick to indexing S since it reminds us this is a state space.

- The set $\mathcal{M}_F[n \times m]$ will denote the set of all matrices of size $n \times m$ with elements in the field F. This may have been alternately seen as $\mathcal{M}_{m \times n}(F)$ but since we almost always assume our probabilities lie on $\mathbb{R}$, it is nice to put focus on our matrix dimensions.

Markov Chains

(Definition) Markov Chain. Say a set of random variables X_i each take a value in a set, called the state space, $S_M = \{1, 2, \ldots, M\}$. Then, a sequence of such random variables $X_0, X_1, \ldots, X_n$ is called a Markov Chain if the following conditions are satisifed:

- $\forall X_i : X_i$ has support and range $S_M = \{1, 2, ..., M\}$.

- *(Markov Property)* The transition probability from state $i \to j$, $\mathrm{P}(X_{n+1} = j \,|\, X_n = i)$ is conditionally independent from all past events in the sequence $X_{n-1} = i', X_{n-2} = i'', \ldots, X_0 = i^{(n-1)}$, excluding the present/last event in the sequence. In other words, given the present, the past and the future are conditionally independent.

$$\forall i, j \in S_M : \mathrm{P}(X_{n+1} = j \,|\, X_n = i) = \mathrm{P}(X_{n+1} = j \,|\, X_n = i, X_{n-1} = i', \ldots, X_0 = i^{(n-1)})$$

(Definition) Transition Matrix. Let $X_0, X_1, \ldots, X_n$ be a Markov Chain with state space S_M. Letting $q_{ij} = \mathrm{P}(X_{n+1} = j \,|\, X_n = i)$ be the transition probability from $i \to j$, then the matrix $Q \in \mathcal{M}_{\mathbb{R}^+}[M \times M]$: $Q = (q_{ij})$ is the *transition matrix* of the chain. Q must satisfy the following conditions to be a valid transition matrix:

- Q is a non-negative matrix. That is, note that $Q \in \mathcal{M}_{\mathbb{R}^+}[M \times M]$ so every $q_{ij} \in \mathbb{R}^+$. This follows because probabilities are necessarily non-negative values.

- The entries of every row i of Q must sum up to 1. This may be understood as applying the law of total probability to the event of transitioning from any given state $\forall i \in S_M$. In other words, the chain has to go somewhere with probability 1.

$$\forall i \in S_M : \sum_{j \in S_M} q_{ij} = 1$$

- Note, it is NOT necessary that the converse holds. The columns of our transition matrix need not sum to 1 for it to be a valid transition matrix.

(Definition) n-step transition probablity. The $n-$step transition probability of $i \to j$ is the probability of being at j exactly n steps after being at i. We denote this value $q_{ij}^{(n)}$:

$$q_{ij}^{(n)} : \mathrm{P}(X_n = j \mid X_0 = i)$$

Realize:

$$q_{ij}^{(2)} = \sum_{k \in S_M} q_{ik} \cdot q_{kj}$$

Because by definition, a Markov Chain is closed under a support/range of S_M so the event $i \to j$ may have taken any intermediate step $k \in S_M$. Realize by notational equivalence, $Q^2 = (q_{ij}^{(2)})$. Inducting over n, we then obtain that:

$$q_{ij}^{(n)} \text{ is the } (i,j) \text{ entry of } Q^n$$

(Definition) Marginal distribution of $\mathbf{X_n}$. Let $\mathbf{t} = (t_1, t_2, \ldots, t_M)$ such that $\forall i \in S_M : t_i = \mathrm{P}(X_0 = i)$. So, $\mathbf{t} \in \mathcal{M}_{\mathbb{R}}[1, M]$. Then, the marginal distribution of X_n is given by the product of the vector $\mathbf{t}Q^n \in \mathcal{M}_{\mathbb{R}}[1, M]$. That is, the jth component of that vector is $P(X_n = j)$ for any $j \in S_M$. We may call $\mathbf{t}$ an initial state distribution.

Classification of states

- A state $i \in S_M$ is said to be **recurrent** if starting from i, the probability is 1 that the chain will *eventually* return to i. If the chain is not recurrent, it is **transient**, meaning that if it starts at i, there is a non-zero probability that it never returns to i.

- Caveat: As we let $n \to \infty$, our Markov chain will gurantee that all transient states will be left forever, no matter how small the probability is. This can be proven by letting the probability be some ϵ, then realizing that by the support of $\mathrm{Geom}(\epsilon)$ is always some finite value, then the equivalence between the Markov property and independent Geometric trials gurantees the existence of some finite value such that there is a success of never returning to i.

(Definition) Reducibility. A Markov chain is said to be **irreducible** if for any $i, j \in S_M$, it is possible to go from $i \to j$ in a finite number of steps with positive probability. In other words:

$$\forall i, j \in S_M : \exists n \in \mathbb{N} : q_{ij}^{(n)} > 0$$

- From our quantifier formulation of irreducible Markov chains, note that we can equivalently say that a chain is irreducible if there is integer $n \in \mathbb{N}$ such that the (i,j) entry of Q^n is positive for any i, j.

- A Markov chain is **reducible** if it is not **irreducible**. Using our quantifier formulation, it means that it suffices to find transient states so that:

$$\exists i, j \in S_M : \nexists n \in \mathbb{N} : q_{ij}^{(n)} > 0$$

(Lemma) Irreducibility $\Rightarrow$ Universal Reccurency. We know that a chain must have atleast one reccurent state. Otherwise, the chain would have nowhere to go and leave all states forever. WLOG, let this reccurent state be state 1. Then, take another state i. Supposing that the chain is irreducible implies there is a finite number of steps, say n, such that there is a positive probability for the chain to go from state $1 \to i$, equal to $q_{1i}^{(n)}$. Since the state 1 is reccurent, it will visit that state infinitely often. Since the transition probability for $1 \to i$ is non-zero, it will also visit state i infinitely often. By induction, the chain visits all states infinitely often, or in other words, all states are reccurent. Hence, irreducibility $\Rightarrow$ all states are reccurent.

(Definition) Periodicity. For a Markov chain with transition matrix P, the period of state i, denoted $d(i)$, is the greatest common divisor of the set of possible return times to i. That is,

$$d(i) = \gcd\{n \in \mathbb{N}^+ : P_{ii}^n > 0\}$$

If $d(i) = 1$, state i is said to be `aperiodic`. If the set of return times is empty, set $d(i) = +\infty$.

Stationary Distributions

(Definition) Stationary Distribution. A row vector $\mathbf{s} \in \mathcal{M}_{\mathbb{R}}[1, M]$, where $\mathbf{s} = (s_1, \ldots, s_M)$ is called a *stationary distribution* for a Markov chain with transition matrix Q if:

$$\forall j \in S_M : \sum_{i \in S_M} s_i q_{ij} = s_j$$

We may rewrite this as a system of linear equations as:

$$\mathbf{s}Q = \mathbf{s}$$

(Theorem) Existence and Uniqueness. If a Markov chain $X_0, X_1, \ldots, X_n$ is irreducible, there exists a unique stationary distribution, say $\pi \in \mathcal{M}_{\mathbb{R}}[1 \times M]$.

Reversibility

(Definition) Reversibility. A Markov chain $X_0, X_1, \ldots, X_n$ is reversible if its transition matrix Q satisfies the reversibility condition for some valid probability distribution $s = (s_1, \ldots, s_M)$:

$$\forall i, j \in S_M : s_i q_{ij} = s_j q_{ji}$$