Discrete time Markov Chains

Thomas Mountford

EPFL

March 1, 2021

MARKOV CHAINS

Definition: A stochastic process is a collection of random variables (or random elements) $\{X_t\}_{t\in T}$ defined on the same probability space. (Usually we take $T = \mathbb{N}$ or $\mathbb{R}$. We will consider Markov chains $\{X_n\}_{n\geq 0}$ taking values a finite or countable state space I. Usually we consider $I = \{1, 2 \cdots N\}$ or $I = \mathbb{Z}$ But not always.

Preparatory Definitions: A *probability vector* λ on I is a a function $\lambda : I \to [0, 1]$ so that $\sum_{I\in I} \lambda_i = 1$.

A matrix p_{ij} on I is a *stochastic matrix* or *transition matrix* if for every $I \in I, \sum_{j\in I} p_{ij} = 1$. For I infinite and P a matrix on I P^2 is not in general defined but it is for stochastic matrices.

Markov Chains: definition

A stochastic process $\{X_n\}_{n\geq 0}$ is a (λ, P) Markov chain on I for λ a probability vector and P a stochastic matrix on I if

(i) $\mathbb{P}(X_0 = i) \;=\; \lambda_i \forall I,$

(ii) For any n and $i_0, i_1, \cdots i_n,\ \mathbb{P}(X_{n+1} = j \mid X_0 = i_0, X_1 = i_1, \cdots X_n = i_n) \;=\; p_{i_n j}$

Condition (ii) is often expressed as X_{n+1} is conditionally independent of $(X_0, X_1, \cdots X_n)$ given $X_n = i_n$ for every $i_n, j \in I$.
We say $\{X_n\}_{n\geq 0}$ is a Markov chain (on I) if it is a (λP) Markov chain forsome λ and P. λ is called the initial distribution (for $\{X_n\}_{n\geq 0}$) and P is the transition matrix.
We can similarly speak of Markov chains $\{X_n\}_{0\leq n\leq N}$

An equivalence

Theorem

$\{X_n\}_{n\geq 0}$ is a (λ, P) Markov chain if and only if $\forall n, i_0, i_1 \cdots i_n \in I$

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

$\{X_n\}_{n\geq 0}$ is a (λ, P) Markov chain, we prove the claimed probability by induction on n. For $n = 0$ it is just the definition of the initial distribution λ. Suppose the claim is true for n, then by (ii)

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots X_n = i_n, X_{n+1} = i_{n+1}) =$$

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, \cdots X_n = i_n)\mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots X_n = i_n)$$

$$= p_{i_n i_{n+1}} \mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots X_n = i_n)$$

$$= p_{i_n i_{n+1}} \quad \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n}$$

Proof continued

Conversely, given the condition applied with $n = 0$, we have that $\forall i,\ \mathbb{P}(X_0 = i) = \lambda_i$. That is condition (i) Equally for condition (ii) $\forall n, i_0, i_1 \cdots i_n \in I$

$$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, \cdots X_n = i_n) =$$

$$\frac{\mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots X_n = i_n, X_{n+1} = i_{n+1})}{\mathbb{P}(X_0 = i_0, X_1 = i_1, \cdots X_n = i_n)} = p_{i_n i_{n+1}}$$

The Markov Property

Theorem

For $\{X_n\}_{n\geq 0}$ *a* (λ, P) *Markov chain, suppose that* $\mathbb{P}(X_m = i) > 0$, *then conditional on event* $\{X_m = i\}$, *the process*

$$Z_n \;=\; X_{m+n} \quad n \geq 0$$

is a (δ_i, P) *Markov chain independent of* $(X_0, X_1, \cdots X_m)$

Define events

- $A \;=\; \{X_0 = i_0, X_1 = i_1, \cdots X_m = i_m(= i)\}$
- $B \;=\; \{X_m = j_0(= i), X_{m+1} = j_1, \cdots X_{m+n} = j_n\}$
- $C \;=\; \{X_m = i)\}$

$$\mathbb{P}(B \mid A \cap C) \;=\; \frac{\mathbb{P}(B \cap A \cap C)}{\mathbb{P}(A \cap C)} \;=\; p_{j_0 j_1} p_{j_1 j_2} p_{j_{n-1} j_n}$$

$=\mathbb{P}(B \mid C) \;=\; \mathbb{P}(Z_0 = j_0, Z_1 = j_1, \cdots Z_n = j_n)$ for (δ_i, P) MC.

λ is not important

So we can apply the Theorem with $m = 0$. For any event A depending on $X_0, X_1, \cdots$, we have by Total Probability

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(X_0 = i)\mathbb{P}(A \mid X_0 = i)$$

$$= \sum_{i \in I} \lambda_i \mathbb{P}_i(A)$$

where $\mathbb{P}_i$ is the probability for a (δ_i, P) Markov chain. That is if we know the probabilities $\mathbb{P}_i$, then we know the probabilities arising out of all (λ, P) Markov chains.

Powers of transition matrices.

Consider a (δ_i, P) Markov chain for fixed i. We know $\mathbb{P}_i(X_1 = j) = p_{ij}$. By the law of total probability

$$\mathbb{P}_i(X_2 = k) = \sum_j \mathbb{P}_i(X_1 = j)\mathbb{P}_i(X_2 = k \mid X_1 = j) =$$

$$\sum_j \mathbb{P}_i(X_1 = j)p_{jk} = \sum_j p_{ij}p_{jk}$$

$= P^2_{ik} = p^2_{ik}$. With an identical argument

$$\mathbb{P}_i(X_3 = k) = \sum_j \mathbb{P}_i(X_12 = j)\mathbb{P}_i(X_3 = k \mid X_12 = j) =$$

$$\sum_j \mathbb{P}_i(X_2 = j)p_{jk} = \sum_j p^2_{ij}p_{jk}$$

$= P^3_{ik}$. For a (λ, P) Markov chain,

$$\mathbb{P}_i(X_r = k) = \sum_i \lambda_i P^r_{ik}$$

Consequences

Theorem

For $\{X_n\}_{n\geq 0}$ *a* (λ, P) *Markov chain, and* r *a strictly positive integer, the process*

$$Z_n \;=\; X_{nr}, \;\; n \geq 0$$

is a (λ, P^r) *Markov chain.*

Theorem

For $\{X_n\}_{n\geq 0}$ *a* (λ, P) *Markov chain, and* $r_1 < r_2 < \cdots < r_k$

$$\mathbb{P}_i(X_{r_1} = i_1, X_{r_2} = i_2, \cdots X_{r_1 k} = i_k) \;=\; \sum_i \lambda_i P^{r_1}_{ii_1} P^{r_2 - r_1}_{i_1 i_2} \cdots P^{r_k - r_{k-1}}_{i_{k-1} i_k}$$