2021s mas651 lecture note

jaemin.oh

March 29, 2021

Contents

Chapter 1

Markov Chain

1.1 Construction, Markov Properties

Let $(S, \mathcal{S})$ be a measurable space. This will be a state space of our Markov Chain.

Definition 1.1.1 (Transition Probability). A function $p : S \times \mathcal{S} \to \mathbb{R}$ is a transition probability if it satisfies two condition:

1. For each $B \in \mathcal{S}$, $p(\cdot, B)$ is a measurable mapping.

2. For each $x \in S$, $p(x, \cdot)$ is a probability measure.

Remark. By the second condition, $p(\cdot, B)$ is a bounded measurable mapping.

If any transition probability is given, then we can define a Markov chain with a natural filtration $\mathcal{F}_n$.

Definition 1.1.2 (Markov Chain). $(X_n, \mathcal{F}_n)$ is a markov chain with transition probability p if
$$P\left(X_{n+1} \in B | \mathcal{F}_n\right) = p\left(X_n, B\right).$$

After defining of Markov chain, we must have a curiosity about the existence. The existence of X_n can be shown by the Kolmogorov extension theorem. First, assume an initial distribution μ on S. Second, define

$$\nu_{0,\dots,n}\left(B_0, \dots B_n\right) = \int_{B_0} \mu(dx_0) \cdots \int_{B_n} p(x_{n-1}, dx_n)$$

Then $\{\nu_{0,\dots n}\}_{n=0}^{\infty}$ is a collection of finite dimensinal distributions which is consistent. Thus, by the Kolmogorov extension theorem, there is a measure P_μ on $(\Omega_0, \mathcal{F}_\infty) = (S^\omega, \mathcal{S}^\omega)$ which satisfies

$$P_\mu\left(X_0 \in B_0, \dots X_n \in B_n\right) = \int_{B_0} \mu(dx_0) \cdots \int_{B_n} p(x_{n-1}, dx_n)$$

where X_n is a coordinate map of $\omega \in \Omega_0$.

Remark. $\omega = (\omega_0, \omega_1, \dots)$ so we can interpret ω_n as a n-th state of our markov chain. $\theta(\omega) = (\omega_n, \omega_{n+1}, \dots)$ is a shift operator, which makes the original n-th state be an initial.

We have constructed P_μ and corresponding X_n. But we don't know whether X_n is a markov chain or not. To show that X_n is actually a markov chain, we should use 1.1.2.

Theorem 1.1.3. X_n above satisfies

$$P_\mu \left(X_{n+1} \in B | \mathcal{F}_n \right) = p \left(X_n, B \right)$$

Proof. To prove this, we must show

$$\int_A 1_{(X_{n+1} \in B)} dP_\mu = \int_A p \left(X_n, B \right) dP_\mu$$

for all $A \in \mathcal{F}_n$. Thanks to $\pi - \lambda$ theorem, we can restrict ourselves to $A = \{X_0 \in B_0, \dots X_n \in B_n\}$. To simplicity, put $B = B_{n+1}$. Then the target equation is equal to

$$P_\mu \left(X_0 \in B_0, \dots, X_{n+1} \in B_{n+1} \right) = \int_{B_0} \mu(dx_0) \cdots \int_{B_n} p(x_{n-1}, dx_n) p(x_n, B_{n+1}).$$

And we want that the last expression is equal to

$$\int_A p(X_n, B) dP_\mu.$$

To do this, first replace $p(x_n, B_{n+1})$ to $1_C(x_n)$. We can easily check that the equality is true for any indicator function. Thus the equality is true for any simple function, and true for any bounded measurable function by BCT. Note that $p(\cdot, B_{n+1})$ is a bounded measurable mapping.

$\square$

To go further, we need a very useful theorem which is called the Monotone Class Theorem.

Theorem 1.1.4 (Monotone Class Theorem)**.** Let $\mathcal{A}$ be a π-system that contains Ω and let $\mathcal{H}$ be a collection of real valued functions which satisfies:

1. If $A \in \mathcal{A}$, then the corresponding indicator function belongs to $\mathcal{H}$.

2. If $f, g \in \mathcal{H}$, then for any real linear combination of them belongs to $\mathcal{H}$.

3. If $0 \leq f_n \uparrow f \leq M$ and $f_n \in \mathcal{H}$, then $f \in \mathcal{H}$.

Then $\mathcal{H}$ contains all $\sigma(\mathcal{A})$ measurable functions which are bounded.

Proof. $\mathcal{G} = \{A : 1_A \in \mathcal{H}\}$ is a λ-system containing $\mathcal{A}$. Thus $\sigma(\mathcal{A}) \subset \mathcal{G}$ which means $1_A \in \mathcal{H}$ for all $A \in \sigma(\mathcal{A})$. Then $\mathcal{H}$ contains all simple functions which are $\sigma(\mathcal{A})$ measurable, hence the conclusion follows from the third condition above. $\square$

Remark. A class of bounded measurable function $f : S \to \mathbb{R}$ which satisfies

$$E\left(f(X_{n+1})|\mathcal{F}_n\right) = \int p(X_n, dy)f(y)$$

is actually a monotone class. Because 1.1.2 says the first condition for $\mathcal{A} = \mathcal{S}$, and the others are satisfied by the elementary properties of expectation and integration.

By induction, we can extend this result by

$$E\left[\prod_{m=0}^{n} f_m(X_m)\right] = \int \mu(dx_0)f_0(x_0)\cdots\int p_{n-1}(x_{n-1}dx_n)f_n(x_n)$$

where the subscript for p stands for temporally inhomogeneous transition probability.

Theorem 1.1.5 (The Markov Property)**.** Let $Y : \Omega_0 \to \mathbb{R}$ be any bounded measurable mapping. Then $E_\mu\left(Y \circ \theta_m | \mathcal{F}_m\right) = E_{X_m}\left[Y\right]$.

Proof. First, assume $Y = \prod_{k=0}^{n} g_k$ where g_k are bounded measurable. By using Dynkin's theorem and the previous result, we can prove the Markov property for this case. We can extend the special case to general case by using the monotone class theorem. $\square$

Remark. $\mathcal{F}_m$ is an information accumulated during time 0 to m. Conditioning on this information, any future(θ_m, starting from time m) quantity can be guessed by the formula. It seems somewhat related to the elementary Markov property. Actually this result implies the usual Markov property. Let's see.

Let $Y = 1_{(X_n = j)}$. Then

$$\begin{aligned}
&P\left(X_n = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\right) \\
&= \frac{P\left(X_0 = i_0, \ldots, X_n = j\right)}{P\left(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\right)} \\
&= \frac{\int_{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}} P\left(X_n = j | \mathcal{F}_{n-1}\right) dP}{P\left(X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\right)} \\
&= \frac{\int_{X_0 = i_0, \ldots, X_{n-1} = i_{n-1}} E\left(Y \circ \theta_{n-1} | \mathcal{F}_{n-1}\right)}{P(X_0 = i_0, \ldots, X_{n-1} = i_{n-1})} \\
&= P\left(X_n = j | X_{n-1} = i_{n-1}\right).
\end{aligned}$$

Theorem 1.1.6 (Chapman-Kolmogorov)**.** Let X_n be a Markov chain. Then $P_x\left(X_{m+n} = z\right) = \sum_y P_x\left(X_m = y\right) P_y\left(X_n = z\right)$.

Proof. This is a direct consequence of the Markov property. Use this:

$$P_x(X_{m+n} = z) = E_x P_x \left(X_{m+n} = z | \mathcal{F}_m \right).$$

$\square$

Remark. This is intuitively clear. To go z from x in exactly $m + n$ steps, we must pass y in m-th step. When a state of m-th step is given, remaining n steps are independent of the past by the Markov property.

Now recall the optional stopping theorem in the martingale theory. That theorem says us that in certain condition, we can regard a stopping time as an index of our martingale. Analogy in Markov chain will be shown.

Let N be any stopping time. Note that N is a stopping time if and only if $\{N = n\} \in \mathcal{F}_n$.

Definition 1.1.7. $\mathcal{F}_N$ is a collection of set A such that $A \cap \{N = m\} \in \mathcal{F}_m$ for all m. Also, $\theta_N(w)$ is $\theta_n(w)$ when $N = n$ and Δ when $N = \infty$. Here, Δ is an extra point added to Ω_0. Since we will only cover the case $N < \infty$, Δ will make no problem.

Theorem 1.1.8 (The Strong Markov Property)**.** Let $Y_n : \Omega_0 \to \mathbb{R}$ be uniformly bounded measurable function. Then

$$E_\mu \left(Y_N \circ \theta_N | \mathcal{F}_N \right) = E_{X_N} \left[Y_N \right]$$

on $\{N < \infty\}$.

Proof. We can write $1_{(N<\infty)} = \sum_{n=0}^{\infty} 1_{(N=n)}$. Let $A \in \mathcal{F}_N$. Then it can be easily shown that

$$E_\mu \left[Y_N \circ \theta_N ; A \cap \{N < \infty\} \right] = E_\mu \left[E_{X_N} \left[Y_N \right] ; A \cap \{N < \infty\} \right].$$

$\square$

Remark. On $\{N < \infty\}$ can be thought of as multiplying both sides by $1_{(N<\infty)}$.

An application of the strong Markov property is:

Theorem 1.1.9. $P_x \left(T_y^k < \infty \right) = \rho_{xy} \rho_{yy}^{k-1}$ for $k \geq 1$.

Proof. Let $Y = 1$ if $w_n = y$ for some n and 0 otherwise. Take $N = T_y^{k-1}$. Then $Y \circ N = 1_{\left(T_y^k < \infty \right)}$. Use this, the strong Markov property, and the induction. $\square$

Theorem 1.1.10 (Reflection Principle)**.** Let $\{\xi_i\}_{i=1}^{\infty}$ be an iid sequence of random variables, with symmetric distribution about 0. Let $S_n = \sum_{i=1}^{n} \xi_i$. Then S_n is clearly Markov chain. If $a > 0$, then

$$P \left(\sup_{m \leq n} S_m \geq a \right) \leq 2P \left(S_n \geq a \right).$$

Proof. temporally omitted. □

¡++¿

Remark. This is a kind of maximal inequality. The maximal quantity of the random variable during time n can be estimated by the quantity of the random variable at time n.