Ma140a Probability
Markov Chains

Juan Pablo Vigneaux

Caltech

Department of Mathematics
California Institute of Technology
Pasadena, CA, USA
vigneaux@caltech.edu

August 3, 2024

Theorem (Monotone Class Theorem, Williams 3.14)

Let $\mathcal{H}$ be a class of bounded functions from a set S into $\mathbb{R}$ satisfying the conditions:

1. *$\mathcal{H}$ is a vector space over $\mathbb{R}$,*
2. *the constant function 1 is an element of $\mathcal{H}$,*
3. *if (f_n) is a sequence of nonnegative bounded functions in $\mathcal{H}$ such that $f_n \uparrow f$ where f is bounded, then $f \in \mathcal{H}$.*

Then if $\mathcal{H}$ contains the indicator function of every set in some π-system $\mathcal{I}$, then $\mathcal{H}$ contains every bounded $\sigma(I)$-measurable function on S.

An important application; cf. Ex. 6 in Workshop 6

Theorem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability triple;* $\mathcal{G}$ *be a sub-*σ*-algebra of* $\mathcal{F}$*;* X *and* Y *be RV taking values in* $(E_X, \mathcal{F}_X)$ *and* $(E_Y, \mathcal{F}_Y)$ *respectively. If* X *is* $\mathcal{G}$*-measurable and* Y *is independent of* $\mathcal{G}$*, then for every bounded function* $\phi : E_X \times E_Y \to \mathbb{R}$*,*

$$\mathbb{E}(\phi(X, Y)|\mathcal{F}) = g(X), \tag{1}$$

where $g(x) = \mathbb{E}(\phi(x, Y))$*.*

Proof.

Let $\mathcal{H}$ be the class of functions for which (1) holds; the class satisfies the properties (1)-(3) in the theorem (follows from the properties of conditional expectations, including the bounded convergence property).
Let $\mathcal{I}$ be the π-system of rectangles $A \times B$ with $A \in \mathcal{F}_X$ and $B \in \mathcal{F}_Y$; we'd like to show that $I_{A\times B} \in \mathcal{H}$, to conclude that $\mathcal{H}$ contains all bounded real-valued measurable functions on $(E_X \times E_Y, \mathcal{F}_X \otimes \mathcal{F}_Y)$.
So remark that if $\phi(x, y) = I_{A\times B}(x, y)$, then for any $G \in \mathcal{G}$,

$$\begin{aligned}\mathbb{E}(\phi(X,Y);G) &= \mathbb{P}(\{X \in A\} \cap \{Y \in B\} \cap G) = \mathbb{P}(\{X \in A\} \cap G)\mathbb{P}(Y \in B)\\ &= \mathbb{E}(I_A(X)\mathbb{P}(Y \in B); G) = \mathbb{E}(g(X); G).\end{aligned}$$

We have used $g(x) = \mathbb{E}(I_A(x)I_B(Y)) = I_A(x)\mathbb{P}(Y \in B)$. □

Definitions

Definition

Let $(S, \mathcal{S})$ be a measurable space. A function $P : S \times \mathcal{S} \to \mathbb{R}$ is said to be a *transition kernel* if:

1. For each $x \in \mathcal{S}$, $A \mapsto p(x, A)$ is a probability measure on $(S, \mathcal{S})$.
2. For each $A \in \mathcal{S}$, $x \mapsto p(x, A)$ is a measurable function.

Definition

Let $(\Omega, \mathcal{F}, (\mathcal{F}_n), \mathbb{P})$ be a filtered space. A process $X = (X_n)_{n \geq 0}$ is a $(\mathcal{F}_n)$-*Markov chain* (MC) with transition kernel p if

$$\forall n \geq 0, \forall B \in \mathcal{S}, \quad \mathbb{P}(X_{n+1} \in B | \mathcal{F}_n) = p(X_n, B). \tag{2}$$

We refer to the law of X_0 as the initial distribution.

Example: Random walk

Let $(Z_n)_{n\geq 1}$ be an i.i.d. sequence of $\mathbb{R}^d$-valued random variables with law μ, and $\mathcal{F}_n = \sigma(Z_1, ..., Z_n)$. For some $z_0 \in \mathbb{R}^d$, define $X_0 = z_0$ a.s. and $X_n = z_0 + Z_1 + \cdots + Z_n$.

Then

$$\mathbb{P}(X_{n+1} \in B|\mathcal{F}_n) = \mathbb{P}(X_n + Z_{n+1} \in B|\mathcal{F}_n) = \mathbb{E}(I_B(X_n + Z_{n+1})|\mathcal{F}_n).$$

Because of Ex. 6 in Workshop 6: introduce $g(x) = \mathbb{E}(I_B(x + Z_{n+1})) = \int I_B(x+z)\,\mathrm{d}\mu(z) = \mu(B-x)$, which is measurable by Fubini's theorem. Since X_n is $\mathcal{F}_n$-measurable and Z_{n+1} is independent of $\mathcal{F}_n$, the r.v. $g(X_n) = \mu(B - X_n)$ is a version of the conditional expectation $\mathbb{E}(I_B(X_n + Z_{n+1})|\mathcal{F}_n)$.

We conclude that X is a MC with transition kernel $p(x, B) = \mu(B - x)$.

(We have used the notation $B - x := \{\, y - x \,:\, y \in B \,\}$.)

Example: Countable state space

Let S be countable, $\mathcal{S} = 2^S$, μ a probability measure on S and $p : S \times S \to [0, 1]$ such that

$$\forall s \in S, \quad \sum_{t \in S} p(s, t) = 1.$$

This extends to sets: $p(s, B) = \sum_{t \in B} p(s, t)$, defining for each s a probability distribution on 2^S.

We consider a collection of S-valued random variables $\{ Z_s^{(n)} : n \geq 0, s \in S \}$ such that for each n and s, $Z_s^{(n)}$ has law $p(s, \cdot)$.

Let $X_0 \sim \mu$, $\mathcal{F}_n = \sigma(X_0, (Z_s^{(1)})_{s \in S}, ..., (Z_s^{(n)})_{s \in S})$ and define $X_{n+1} = Z_{X_n}^{(n+1)}$. The resulting process $X = (X_n)_{n \geq 0}$ is a MC with transition kernel p.

Proof: for any $B \in \mathcal{S}$, $\mathbb{P}(X_{n+1} \in B | \mathcal{F}_n) = \mathbb{E}(I_B(Z_{X_n}^{(n+1)}) | \mathcal{F}_n)$. As before, one introduces $g(x) = \mathbb{E}(I_B(Z_x^{(n+1)}) = p(x, B)$, and concludes by noting that $g(X_n)$ is a version of $\mathbb{E}(I_B(Z_{X_n}^{(n+1)}) | \mathcal{F}_n)$.

One-step expectations

Theorem

Let (X_n) be a MC on S with transition kernel p and initial distribution μ. Then for any bounded measurable function $f : S \to \mathbb{R}$,

$$\mathbb{E}(f(X_{n+1})|\mathcal{F}_n) = \int f(y)p(X_n, \mathrm{d}y). \tag{3}$$

The result follows directly from the definition of MC using the monotone class theorem. (*Let $\mathcal{H}$ be the class of functions for which* (3) *holds...*)

Markov property

Theorem

Let $(X_n)_{n\geq 0}$ *be a MC with transition kernel* p. *Let* h *be a bounded measurable function from* $S^{\mathbb{N}}$ *to* $\mathbb{R}$. *Then*

$$\mathbb{E}(h(X_n, X_{n+1}, ...)|\mathcal{F}_n) = \phi(X_n) \tag{4}$$

where $\phi(x) = \mathbb{E}_x(h)$, *the expectation of* h *evaluated on a MC* $(Y_n)_{n\geq 0}$ *starting at* $x \in S$ *and with transitions* p.