

Stat 709: Mathematical Statistics

Lecture 10

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Lecture 10: Markov chains

Markov chain

An important example of dependent sequence of random variables in statistical application

A sequence of random vectors $\{X_n : n = 1, 2, \dots\}$ is a *Markov chain* or *Markov process* iff

$$P(B|X_1, \dots, X_n) = P(B|X_n) \text{ a.s.}, \quad B \in \sigma(X_{n+1}), \quad n = 2, 3, \dots$$

We call the previous equation the “Markov property”.

Remarks

- X_{n+1} (tomorrow) is conditionally independent of (X_1, \dots, X_{n-1}) (the past), given X_n (today).
- (X_1, \dots, X_{n-1}) is not necessarily independent of (X_n, X_{n+1}) .
- A sequence of independent random vectors forms a Markov chain

Lecture 10: Markov chains

Markov chain

An important example of dependent sequence of random variables in statistical application

A sequence of random vectors $\{X_n : n = 1, 2, \dots\}$ is a *Markov chain* or *Markov process* iff

$$P(B|X_1, \dots, X_n) = P(B|X_n) \text{ a.s.}, \quad B \in \sigma(X_{n+1}), \quad n = 2, 3, \dots$$

We call the previous equation the “Markov property”.

Remarks

- X_{n+1} (tomorrow) is conditionally independent of (X_1, \dots, X_{n-1}) (the past), given X_n (today).
- (X_1, \dots, X_{n-1}) is not necessarily independent of (X_n, X_{n+1}) .
- A sequence of independent random vectors forms a Markov chain

Example 1.24 (First-order autoregressive processes)

Let $\varepsilon_1, \varepsilon_2, \dots$ be independent random variables defined on a probability space, $X_1 = \varepsilon_1$, and $X_{n+1} = \rho X_n + \varepsilon_{n+1}$, $n = 1, 2, \dots$, where ρ is a constant in \mathcal{R} .

Then $\{X_n\}$ is called a first-order autoregressive process.

We now show that $\{X_n\}$ is a Markov chain

We need to show the Markov property, i.e., for any $B \in \mathcal{B}$ and $n = 1, 2, \dots$,

$$P(X_{n+1} \in B | X_1, \dots, X_n) = P_{\varepsilon_{n+1}}(B - \rho X_n) = P(X_{n+1} \in B | X_n) \text{ a.s.,}$$

where $B - y = \{x \in \mathcal{R} : x + y \in B\}$.

For any $y \in \mathcal{R}$,

$$P_{\varepsilon_{n+1}}(B - y) = P(\varepsilon_{n+1} + y \in B) = \int I_B(x + y) dP_{\varepsilon_{n+1}}(x)$$

and, by Fubini's theorem, $P_{\varepsilon_{n+1}}(B - y)$ is Borel.

Hence, $P_{\varepsilon_{n+1}}(B - \rho X_n)$ is Borel w.r.t. $\sigma(X_n)$ and, thus, is Borel w.r.t. $\sigma(X_1, \dots, X_n)$.

Example 1.24 (First-order autoregressive processes)

Let $\varepsilon_1, \varepsilon_2, \dots$ be independent random variables defined on a probability space, $X_1 = \varepsilon_1$, and $X_{n+1} = \rho X_n + \varepsilon_{n+1}$, $n = 1, 2, \dots$, where ρ is a constant in \mathcal{R} .

Then $\{X_n\}$ is called a first-order autoregressive process.

We now show that $\{X_n\}$ is a Markov chain

We need to show the Markov property, i.e., for any $B \in \mathcal{B}$ and $n = 1, 2, \dots$,

$$P(X_{n+1} \in B | X_1, \dots, X_n) = P_{\varepsilon_{n+1}}(B - \rho X_n) = P(X_{n+1} \in B | X_n) \text{ a.s.,}$$

where $B - y = \{x \in \mathcal{R} : x + y \in B\}$.

For any $y \in \mathcal{R}$,

$$P_{\varepsilon_{n+1}}(B - y) = P(\varepsilon_{n+1} + y \in B) = \int I_B(x + y) dP_{\varepsilon_{n+1}}(x)$$

and, by Fubini's theorem, $P_{\varepsilon_{n+1}}(B - y)$ is Borel.

Hence, $P_{\varepsilon_{n+1}}(B - \rho X_n)$ is Borel w.r.t. $\sigma(X_n)$ and, thus, is Borel w.r.t. $\sigma(X_1, \dots, X_n)$.

Example 1.24 (First-order autoregressive processes)

Let $\varepsilon_1, \varepsilon_2, \dots$ be independent random variables defined on a probability space, $X_1 = \varepsilon_1$, and $X_{n+1} = \rho X_n + \varepsilon_{n+1}$, $n = 1, 2, \dots$, where ρ is a constant in \mathcal{R} .

Then $\{X_n\}$ is called a first-order autoregressive process.

We now show that $\{X_n\}$ is a Markov chain

We need to show the Markov property, i.e., for any $B \in \mathcal{B}$ and $n = 1, 2, \dots$,

$$P(X_{n+1} \in B | X_1, \dots, X_n) = P_{\varepsilon_{n+1}}(B - \rho X_n) = P(X_{n+1} \in B | X_n) \text{ a.s.,}$$

where $B - y = \{x \in \mathcal{R} : x + y \in B\}$.

For any $y \in \mathcal{R}$,

$$P_{\varepsilon_{n+1}}(B - y) = P(\varepsilon_{n+1} + y \in B) = \int I_B(x + y) dP_{\varepsilon_{n+1}}(x)$$

and, by Fubini's theorem, $P_{\varepsilon_{n+1}}(B - y)$ is Borel.

Hence, $P_{\varepsilon_{n+1}}(B - \rho X_n)$ is Borel w.r.t. $\sigma(X_n)$ and, thus, is Borel w.r.t. $\sigma(X_1, \dots, X_n)$.

Example 1.24 (First-order autoregressive processes)

Let $\varepsilon_1, \varepsilon_2, \dots$ be independent random variables defined on a probability space, $X_1 = \varepsilon_1$, and $X_{n+1} = \rho X_n + \varepsilon_{n+1}$, $n = 1, 2, \dots$, where ρ is a constant in \mathcal{R} .

Then $\{X_n\}$ is called a first-order autoregressive process.

We now show that $\{X_n\}$ is a Markov chain

We need to show the Markov property, i.e., for any $B \in \mathcal{B}$ and $n = 1, 2, \dots$,

$$P(X_{n+1} \in B | X_1, \dots, X_n) = P_{\varepsilon_{n+1}}(B - \rho X_n) = P(X_{n+1} \in B | X_n) \text{ a.s.},$$

where $B - y = \{x \in \mathcal{R} : x + y \in B\}$.

For any $y \in \mathcal{R}$,

$$P_{\varepsilon_{n+1}}(B - y) = P(\varepsilon_{n+1} + y \in B) = \int I_B(x + y) dP_{\varepsilon_{n+1}}(x)$$

and, by Fubini's theorem, $P_{\varepsilon_{n+1}}(B - y)$ is Borel.

Hence, $P_{\varepsilon_{n+1}}(B - \rho X_n)$ is Borel w.r.t. $\sigma(X_n)$ and, thus, is Borel w.r.t. $\sigma(X_1, \dots, X_n)$.

Example 1.24 (continued)

Let $B_j \in \mathcal{B}$, $j = 1, \dots, n$, and $A = \bigcap_{j=1}^n X_j^{-1}(B_j)$.

Since $\varepsilon_{n+1} + \rho X_n = X_{n+1}$ and ε_{n+1} is independent of (X_1, \dots, X_n) , it follows from Theorem 1.2 and Fubini's theorem that

$$\begin{aligned} \int_A P_{\varepsilon_{n+1}}(B - \rho X_n) dP &= \int_{x_j \in B_j, j=1, \dots, n} \int_{t \in B - \rho x_n} dP_{\varepsilon_{n+1}}(t) dP_X(x) \\ &= \int_{x_j \in B_j, j=1, \dots, n, x_{n+1} \in B} dP_{(X, \varepsilon_{n+1})}(x, t) \\ &= P\left(A \cap X_{n+1}^{-1}(B)\right), \end{aligned}$$

where X and x denote (X_1, \dots, X_n) and (x_1, \dots, x_n) , respectively, and x_{n+1} denotes $\rho x_n + t$.

Using this and the argument in the end of the proof for Proposition 1.11, we obtain $P(X_{n+1} \in B | X_1, \dots, X_n) = P_{\varepsilon_{n+1}}(B - \rho X_n)$ a.s.

The proof for $P_{\varepsilon_{n+1}}(B - \rho X_n) = P(X_{n+1} \in B | X_n)$ a.s. is similar and simpler.

Proposition 1.12 (Characterizations of Markov chains)

A sequence of random vectors $\{X_n\}$ is a Markov chain if and only if one of the following three conditions holds.

- (a) For any $n = 2, 3, \dots$ and any integrable $h(X_{n+1})$ with a Borel function h ,

$$E[h(X_{n+1})|X_1, \dots, X_n] = E[h(X_{n+1})|X_n] \text{ a.s.}$$

- (b) For any $n = 1, 2, \dots$ and $B \in \sigma(X_{n+1}, X_{n+2}, \dots)$,

$$P(B|X_1, \dots, X_n) = P(B|X_n) \text{ a.s.}$$

("the past and the future are conditionally independent given the present")

- (c) For any $n = 2, 3, \dots$, $A \in \sigma(X_1, \dots, X_n)$, and $B \in \sigma(X_{n+1}, X_{n+2}, \dots)$,

$$P(A \cap B|X_n) = P(A|X_n)P(B|X_n) \text{ a.s.}$$

Proof

(i) The equivalence between (a) and the Markov property.

It is clear that (a) implies the Markov property.

If h is a simple function, then the Markov property and Proposition 1.10(iii) imply (a).

If h is nonnegative, then there are nonnegative simple functions $h_1 \leq h_2 \leq \dots \leq h$ such that $h_j \rightarrow h$.

Then the Markov property together with Proposition 1.10(iii) and (x) imply (a).

Since $h = h_+ - h_-$, we conclude that the Markov property implies (a).

(ii) The equivalence between (b) and the Markov property.

It is clear that (b) implies the Markov property.

Note that $\sigma(X_{n+1}, X_{n+2}, \dots) = \sigma\left(\bigcup_{j=1}^{\infty} \sigma(X_{n+1}, \dots, X_{n+j})\right)$ (Exercise 19).

Hence, to show that the Markov property implies (b), it suffices to show that $P(B|X_1, \dots, X_n) = P(B|X_n)$ a.s. for $B \in \sigma(X_{n+1}, \dots, X_{n+j})$ for any $j = 1, 2, \dots$

We use induction.

The result for $j = 1$ follows from the Markov property.

(i) The equivalence between (a) and the Markov property.

It is clear that (a) implies the Markov property.

If h is a simple function, then the Markov property and Proposition 1.10(iii) imply (a).

If h is nonnegative, then there are nonnegative simple functions $h_1 \leq h_2 \leq \dots \leq h$ such that $h_j \rightarrow h$.

Then the Markov property together with Proposition 1.10(iii) and (x) imply (a).

Since $h = h_+ - h_-$, we conclude that the Markov property implies (a).

(ii) The equivalence between (b) and the Markov property.

It is clear that (b) implies the Markov property.

Note that $\sigma(X_{n+1}, X_{n+2}, \dots) = \sigma\left(\bigcup_{j=1}^{\infty} \sigma(X_{n+1}, \dots, X_{n+j})\right)$ (Exercise 19).

Hence, to show that the Markov property implies (b), it suffices to show that $P(B|X_1, \dots, X_n) = P(B|X_n)$ a.s. for $B \in \sigma(X_{n+1}, \dots, X_{n+j})$ for any $j = 1, 2, \dots$

We use induction.

The result for $j = 1$ follows from the Markov property.

Proof (continued)

Suppose that the result holds for any $B \in \sigma(X_{n+1}, \dots, X_{n+j})$.

To show the result for any $B \in \sigma(X_{n+1}, \dots, X_{n+j+1})$, it is enough (why?) to show that for any $B_1 \in \sigma(X_{n+j+1})$ and any $B_2 \in \sigma(X_{n+1}, \dots, X_{n+j})$, $P(B_1 \cap B_2 | X_1, \dots, X_n) = P(B_1 \cap B_2 | X_n)$ a.s.

From the proof in (i), the induction assumption implies

$$E[h(X_{n+1}, \dots, X_{n+j}) | X_1, \dots, X_n] = E[h(X_{n+1}, \dots, X_{n+j}) | X_n] \quad (1)$$

for any Borel function h .

The result follows from

$$\begin{aligned} E(I_{B_1} I_{B_2} | X_1, \dots, X_n) &= E[E(I_{B_1} I_{B_2} | X_1, \dots, X_{n+j}) | X_1, \dots, X_n] \\ &= E[I_{B_2} E(I_{B_1} | X_1, \dots, X_{n+j}) | X_1, \dots, X_n] \\ &= E[I_{B_2} E(I_{B_1} | X_{n+j}) | X_1, \dots, X_n] \\ &= E[I_{B_2} E(I_{B_1} | X_{n+j}) | X_n] \\ &= E[I_{B_2} E(I_{B_1} | X_n, \dots, X_{n+j}) | X_n] \\ &= E[E(I_{B_1} I_{B_2} | X_n, \dots, X_{n+j}) | X_n] \\ &= E(I_{B_1} I_{B_2} | X_n) \text{ a.s.,} \end{aligned}$$

Proof (continued)

where the first and last equalities follow from Proposition 1.10(v), the second and sixth equalities follow from Proposition 1.10(vi), the third and fifth equalities follow from the Markov property, and the fourth equality follows from (1).

(iii) The equivalence between (b) and (c)

Let $A \in \sigma(X_1, \dots, X_n)$ and $B \in \sigma(X_{n+1}, X_{n+2}, \dots)$.

If (b) holds, then

$$\begin{aligned} E(I_A I_B | X_n) &= E[E(I_A I_B | X_1, \dots, X_n) | X_n] \\ &= E[I_A E(I_B | X_1, \dots, X_n) | X_n] \\ &= E[I_A E(I_B | X_n) | X_n] \\ &= E(I_A | X_n) E(I_B | X_n), \end{aligned}$$

which is (c).

Proof (continued)

where the first and last equalities follow from Proposition 1.10(v), the second and sixth equalities follow from Proposition 1.10(vi), the third and fifth equalities follow from the Markov property, and the fourth equality follows from (1).

(iii) The equivalence between (b) and (c)

Let $A \in \sigma(X_1, \dots, X_n)$ and $B \in \sigma(X_{n+1}, X_{n+2}, \dots)$.

If (b) holds, then

$$\begin{aligned} E(I_A I_B | X_n) &= E[E(I_A I_B | X_1, \dots, X_n) | X_n] \\ &= E[I_A E(I_B | X_1, \dots, X_n) | X_n] \\ &= E[I_A E(I_B | X_n) | X_n] \\ &= E(I_A | X_n) E(I_B | X_n), \end{aligned}$$

which is (c).

Proof (continued)

Assume that (c) holds.

Let $A_1 \in \sigma(X_n)$, $A_2 \in \sigma(X_1, \dots, X_{n-1})$, and $B \in \sigma(X_{n+1}, X_{n+2}, \dots)$.

Then

$$\begin{aligned} \int_{A_1 \cap A_2} E(I_B | X_n) dP &= \int_{A_1} I_{A_2} E(I_B | X_n) dP \\ &= \int_{A_1} E[I_{A_2} E(I_B | X_n) | X_n] dP \\ &= \int_{A_1} E(I_{A_2} | X_n) E(I_B | X_n) dP \\ &= \int_{A_1} E(I_{A_2} I_B | X_n) dP \\ &= P(A_1 \cap A_2 \cap B). \end{aligned}$$

Since disjoint unions of events of the form $A_1 \cap A_2$ as specified above generate $\sigma(X_1, \dots, X_n)$, this shows that $E(I_B | X_n) = E(I_B | X_1, \dots, X_n)$ a.s., which is (b).