632 Introduction to Stochastic Processes

Markov Chain Supplement

Let S be a countable state space, and $\mathbf{P} = \{p(x, y)\}_{x,y \in S}$ a transition matrix on S. Let $\Omega = S^{\mathbf{Z}_+}$ be the space of sequences $\omega = (x_0, x_1, x_2, \dots)$ with entries from S. On Ω we define the coordinate random variables X_n by $X_n(\omega) = x_n$ for $n = 0, 1, 2, \dots$

The general construction theory of stochastic processes assures that for every choice of initial distribution μ on S there exists a probability measure P_μ on Ω with the property that

$$\begin{aligned} P_\mu\{X_0 = x_0, X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \\ = \mu(x_0)p(x_0, x_1)p(x_1, x_2)\cdots p(x_{n-1}, x_n) \end{aligned} \tag{1}$$

for any choice of states $x_0, x_1, x_2, \dots, x_n \in S$. Under this probability measure, the process $\{X_n\}$ is the *Markov chain with initial distribution* μ *and transition matrix* $\mathbf{P}$.

When μ is concentrated on a state x (the initial state X_0 is not really random but equals x), then P_μ is denoted by P_x.

Given a stochastic process $\{X_n\}$ the basic *Markov property* is usually stated as

$$P_\mu\big[X_{n+1} = x_{n+1} \,\big|\, X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0\big] = p(x_n, x_{n+1}). \tag{2}$$

This is required to hold whenever the conditioning event has positive probability. It is true for the process $\{X_n\}$ constructed above, no matter what the initial distribution is. Note that we are discussing here the case of a *time-homogeneous* Markov chain, where the transition probability does not change with time.

We state here some extensions of (2) that are useful for computations. They all embody the idea that if we know the present state, then the past does not influence probabilities of the future, and each time the Markov chain restarts itself anew, using the current state as the new initial state. The most general form of this statement is Theorem 1 below. The lemmas lead up to it by establishing successively more general statements.

After Theorem 1 we establish a result in the opposite direction: that an arbitrary process with the Markov property (2) satisfies the multiplicative formula (1). This is Theorem 2.

Lemma 1 *Assume* (1). *For any states* $y_0, \dots, y_n$, $x_1, \dots, x_m$ *from* S, *and any initial distribution* μ,

$$\begin{aligned} P_\mu\big[X_{n+m} = x_m, \dots, X_{n+2} = x_2, X_{n+1} = x_1 \,|\, X_n = y_n, X_{n-1} = y_{n-1}, \dots, X_0 = y_0\big] \\ = p(y_n, x_1)p(x_1, x_2)\cdots p(x_{m-1}, x_m) \\ = P_{y_n}\big[X_1 = x_1, X_2 = x_2, \dots, X_m = x_m\big]. \end{aligned} \tag{3}$$

provided the conditioning event has positive probability.

Proof. We compute using property (1):

$$
\begin{aligned}
&P_\mu\big[X_{n+m}=x_m,\ldots,X_{n+2}=x_2,X_{n+1}=x_1 \,|\, X_n=y_n,X_{n-1}=y_{n-1},\ldots,X_0=y_0\big]\\
&\quad=\frac{P_\mu\big[X_{n+m}=x_m,\ldots,X_{n+1}=x_1,X_n=y_n,\ldots,X_0=y_0\big]}{P_\mu\big[X_n=y_n,\ldots,X_0=y_0\big]}\\
&\quad=\frac{\mu(y_0)p(y_0,y_1)\cdots p(y_{n-1},y_n)p(y_n,x_1)\cdots p(x_{m-1},x_m)}{\mu(y_0)p(y_0,y_1)\cdots p(y_{n-1},y_n)}\\
&\quad=p(y_n,x_1)p(x_1,x_2)\cdots p(x_{m-1},x_m)\\
&\quad=P_{y_n}\big[X_1=x_1,X_2=x_2,\ldots,X_m=x_m\big].
\end{aligned}
$$

The second last equality follows from cancelling, the last equality by property (1). ∎

Eq. (3) remains valid if we add a condition on X_n inside the probability:

$$
\begin{aligned}
&P_\mu\big[X_{n+m}=x_m,\ldots,X_{n+1}=x_1,X_n=x_0 \,|\, X_n=y_n,X_{n-1}=y_{n-1},\ldots,X_0=y_0\big]\\
&\quad=P_{y_n}\big[X_0=x_0,X_1=x_1,\ldots,X_m=x_m\big].
\end{aligned}
\tag{4}
$$

If $x_0=y_n$ then adding or removing the condition $X_n=x_0$ does not affect the value of the first probability above, and adding or removing the condition $X_0=x_0$ does not affect the value of the second probability. Then (4) reduces to (3). On the other hand, if $x_0\neq y_n$, then both probabilities above are zero because X_n (and X_0) cannot simultaneously equal both x_0 and y_n.

Lemma 2 *Assume* (1). *For any states $y_0,\ldots,y_n$ from S, any subset $U\subseteq S^{m+1}$, and any initial distribution μ,*

$$
\begin{aligned}
&P_\mu\big[(X_n,X_{n+1},\ldots,X_{n+m})\in U \,|\, X_n=y_n,X_{n-1}=y_{n-1},\ldots,X_0=y_0\big]\\
&\quad=P_{y_n}\big[(X_0,X_1,\ldots,X_m)\in U\big].
\end{aligned}
\tag{5}
$$

provided the conditioning event has positive probability.

Proof. This follows from property (4) by addition:

$$
\begin{aligned}
&P_\mu\big[(X_n,\ldots,X_{n+m})\in U \,|\, X_n=y_n,\ldots,X_0=y_0\big]\\
&\quad=\sum_{(x_0,\ldots,x_m)\in U}P_\mu\big[X_{n+m}=x_m,\ldots,X_n=x_0 \,|\, X_n=y_n,\ldots,X_0=y_0\big]\\
&\quad=\sum_{(x_0,\ldots,x_m)\in U}P_{y_n}\big[X_0=x_0,\ldots,X_m=x_m\big]\\
&\quad=P_{y_n}\big[(X_0,\ldots,X_m)\in U\big].
\end{aligned}
$$

∎

Lemma 3 *Assume* (1). *For any states* $y_0, \dots, y_n$ *from* S, *any event* $U \subseteq S^{\mathbf{Z}_+}$, *and any initial distribution* μ,

$$\begin{aligned} &P_\mu\big[(X_n, X_{n+1}, X_{n+2}, \dots) \in U \,|\, X_n = y_n, X_{n-1} = y_{n-1}, \dots, X_0 = y_0\big] \\ &\quad = P_{y_n}\big[(X_0, X_1, X_2, \dots) \in U\big]. \end{aligned} \tag{6}$$

provided the conditioning event has positive probability.

Proof. This follows from (5) with a little leap of faith (let $m \to \infty$). Rigorous justification needs measure theory. ∎

Finally the most general formula where we add an arbitrary event in the condition.

Theorem 1 *Assume* (1). *Let* $x \in S$, $B \subseteq S^{n+1}$ *and* $U \subseteq S^{\mathbf{Z}_+}$. *Then for any initial distribution* μ,

$$\begin{aligned} &P_\mu\big[(X_n, X_{n+1}, X_{n+2}, \dots) \in U \;\big|\; X_n = x, (X_0, \dots, X_n) \in B\big] \\ &\quad = P_x\big[(X_0, X_1, X_2, \dots) \in U\big] \end{aligned} \tag{7}$$

provided the conditioning event has positive probability.

Proof. Let B_0 be the set of $(n+1)$-tuples $(y_0, \dots, y_n)$ in B that satisfy $y_n = x$. Then

$$\big\{X_n = x, (X_0, \dots, X_n) \in B\big\} = \big\{(X_0, \dots, X_n) \in B_0\big\}.$$

Use (6) in the next calculation.

$$\begin{aligned} &P_\mu\big[(X_n, X_{n+1}, X_{n+2}, \dots) \in U \;\big|\; X_n = x, (X_0, \dots, X_n) \in B\big] \\ &= \frac{P_\mu\big[(X_n, X_{n+1}, X_{n+2}, \dots) \in U, X_n = x, (X_0, \dots, X_n) \in B\big]}{P_\mu\big[X_n = x, (X_0, \dots, X_n) \in B\big]} \\ &= \frac{\displaystyle\sum_{(y_0,\dots,y_n) \in B_0} P_\mu\big[(X_n, X_{n+1}, X_{n+2}, \dots) \in U, (X_0, \dots, X_n) = (y_0, \dots, y_n)\big]}{P_\mu\big[X_n = x, (X_0, \dots, X_n) \in B\big]} \\ &= \Bigg\{ \sum_{(y_0,\dots,y_n) \in B_0} P_\mu\big[(X_n, X_{n+1}, X_{n+2}, \dots) \in U \;\big|\; (X_0, \dots, X_n) = (y_0, \dots, y_n)\big] \\ &\qquad\qquad \times P_\mu\big[(X_0, \dots, X_n) = (y_0, \dots, y_n)\big] \Bigg\} \\ &\qquad\qquad\qquad \times \frac{1}{P_\mu\big[X_n = x, (X_0, \dots, X_n) \in B\big]} \end{aligned}$$

$$= \frac{\displaystyle\sum_{(y_0,\dots,y_n)\in B_0} P_x\bigl[(X_0,X_1,X_2,\dots)\in U\bigr]\cdot P_\mu\bigl[(X_0,\dots,X_n)=(y_0,\dots,y_n)\bigr]}{P_\mu\bigl[X_n=x,(X_0,\dots,X_n)\in B\bigr]}$$
$$= \frac{P_x\bigl[(X_0,X_1,X_2,\dots)\in U\bigr]\cdot P_\mu\bigl[X_n=x,(X_0,\dots,X_n)\in B\bigr]}{P_\mu\bigl[X_n=x,(X_0,\dots,X_n)\in B\bigr]}.$$

Cancelling above leaves the right side of (7). ∎

Now for a result in the opposite direction.

Theorem 2 *Let S be a countable state space, and $\mathbf{P}=\{p(i,j)\}_{i,j\in S}$ a transition matrix on S. Let $(X_n : n\in\mathbf{Z}_+)$ be an S-valued stochastic process defined on an arbitrary probability space $(\Omega,\mathcal{F},P)$. Assume*

$$P\bigl[X_{n+1}=x_{n+1}\bigm| X_n=x_n,X_{n-1}=x_{n-1},\dots,X_0=x_0\bigr]=p(x_n,x_{n+1}) \tag{8}$$

for all states $x_0,x_1,x_2,\dots,x_{n+1}\in S$ such that

$$P\bigl[X_0=x_0,X_1=x_1,X_2=x_2,\dots,X_n=x_n\bigr]>0,$$

in other words the conditioning event has positive probability. Then

$$\begin{aligned} &P\bigl\{X_0=x_0,X_1=x_1,X_2=x_2,\dots,X_n=x_n\bigr\}\\ &\qquad=P[X_0=x_0]p(x_0,x_1)p(x_1,x_2)\cdots p(x_{n-1},x_n)\end{aligned} \tag{9}$$

for any choice of states $x_0,x_1,x_2,\dots,x_n\in S$.

Proof. Statement (9) is proved by induction on n. If $n=0$ both sides are $P[X_0=x_0]$. Assume (9) is true for n and any states $x_0,x_1,x_2,\dots,x_n$. Consider the probability

$$P\bigl\{X_0=x_0,X_1=x_1,\dots,X_n=x_n,X_{n+1}=x_{n+1}\bigr\}$$

for some choice of states $x_0,x_1,x_2,\dots,x_{n+1}$. We check the formula (9) for $n+1$ separately for two cases.

Case 1: $P\bigl[X_0=x_0,X_1=x_1,\dots,X_n=x_n\bigr]>0$.
In this case we can condition on the past, then apply (8) and induction to write

$$\begin{aligned} &P\bigl\{X_0=x_0,X_1=x_1,X_2=x_2,\dots,X_n=x_n,X_{n+1}=x_{n+1}\bigr\}\\ &=P\bigl\{X_0=x_0,X_1=x_1,X_2=x_2,\dots,X_n=x_n\bigr\}\\ &\qquad\times P\bigl[X_{n+1}=x_{n+1}\bigm| X_n=x_n,X_{n-1}=x_{n-1},\dots,X_0=x_0\bigr]\\ &=P[X_0=x_0]p(x_0,x_1)p(x_1,x_2)\cdots p(x_{n-1},x_n)\;\times\;p(x_n,x_{n+1}).\end{aligned}$$

Case 2: $P\big[X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\big] = 0$.
By the induction assumption it then also follows that

$$P[X_0 = x_0]p(x_0, x_1)p(x_1, x_2)\cdots p(x_{n-1}, x_n) = 0.$$

The equality

$$\begin{aligned}
&P\big\{X_0 = x_0, X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n, X_{n+1} = x_{n+1}\big\} \\
&= P[X_0 = x_0]p(x_0, x_1)p(x_1, x_2)\cdots p(x_{n-1}, x_n)p(x_n, x_{n+1})
\end{aligned}$$

then follows because both sides equal zero. The left-hand side equals zero because the event that goes up to $n+1$ is a subset of the event that goes up to n, so it also has probability zero. ∎