

APPENDIX

2

Probability

The purpose of this appendix is to provide some information from probability theory that can be used to develop a formal definition of a Markov chain and to prove some results from Chapter 10.

Probability

DEFINITION

For each event E of the sample space S , the **probability** of E (denoted $P(E)$) is a number that has the following three properties:

- $0 \leq P(E) \leq 1$
- $P(S) = 1$
- For any sequence of mutually exclusive events E_1, E_2, \dots ,

$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

Properties of Probability

- $P(\emptyset) = 0$
- $P(E^c) = 1 - P(E)$
- $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- If E and F are mutually exclusive events, $P(E \cup F) = P(E) + P(F)$

DEFINITION

The **conditional probability** of E given F (denoted $P(E|F)$), is the probability that E occurs given that F has occurred, is

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Law of Total Probability

Let F_1, F_2, \dots be a sequence of mutually exclusive events for which

$$\bigcup_{n=1}^{\infty} F_n = S$$

Then for any event E in the sample space S ,

$$P(E) = \sum_{n=1}^{\infty} P(E|F_n)P(F_n)$$

Random Variables and Expectation

DEFINITION

A random variable is a real-valued function defined on the sample space S . A **discrete random variable** is a random variable that takes on at most a countable number of possible values.

Only discrete random variables will be considered in this text; random variables that take on an uncountably infinite set of values are considered in advanced courses in probability theory. In Section 10.3, the expected value of a discrete random variable was defined. The expected value of a discrete random variable may also be defined using a function called the *probability mass function*.

DEFINITION

The **probability mass function** p of a discrete random variable X is the real-valued function defined by $p(a) = P(X = a)$.

DEFINITION

The **expected value** of a discrete random variable X is

$$E[X] = \sum_x xp(x)$$

where the sum is taken over all x with $p(x) > 0$.

Notice that if the random variable takes on the values x_1, x_2, \dots with positive probability, then the expected value of the random variable is

$$\sum_x xp(x) = x_1P(X = x_1) + x_2P(X = x_2) + \dots$$

which matches the definition of expected value given in Section 10.3. Using the definition above, it is straightforward to show that expected value has the following properties.

Properties of Expected Value

For any real constant k and any discrete random variables X and Y ,

1. $E[kX] = kE[X]$
2. $E[X + k] = E[X] + k$
3. $E[X + Y] = E[X] + E[Y]$
4. If f is a real-valued function, then $f(X)$ is a discrete random variable, and $E[f(X)] = \sum_x f(x)p(x)$, where the sum is taken over all x with $p(x) > 0$.

Just as probabilities can be affected by whether an event occurs, so can expected values.

DEFINITION

Let X be a discrete random variable and let F be an event in the sample space S . Then the **conditional expected value** of X given F is

$$E[X|F] = \sum_x xP(X = x|F)$$

where the sum is taken over all x with $p(x) > 0$.

There is a law of total probability for expected value that will be used to prove a result from Chapter 10. Its statement and its proof follow.

Law of Total Probability for Expected Value

Let F_1, F_2, \dots be a sequence of mutually exclusive events for which

$$\bigcup_{n=1}^{\infty} F_n = S$$

Then, for any discrete random variable X ,

$$E[X] = \sum_{n=1}^{\infty} E[X|F_n]P(F_n)$$

PROOF Let F_1, F_2, \dots be a sequence of mutually exclusive events for which $\bigcup_{n=1}^{\infty} F_n = S$, and let X be a discrete random variable. Then, using the definition of expected value and the law of total probability,

$$\begin{aligned} E[X] &= \sum_x xp(x) = \sum_x xP(X = x) \\ &= \sum_x x \sum_{n=1}^{\infty} P(X = x|F_n)P(F_n) \\ &= \sum_{n=1}^{\infty} P(F_n) \sum_x xP(X = x|F_n) \\ &= \sum_{n=1}^{\infty} E[X|F_n]P(F_n) \end{aligned}$$



Markov Chains

In Section 4.9, a Markov chain was defined as a sequence of vectors. In order to understand Markov chains from a probabilistic standpoint, it is better to define a Markov chain as a sequence of random variables. To begin, consider any collection of random variables. This is called a stochastic process.

DEFINITION

A **stochastic process** $\{X_n : n \in T\}$ is a collection of random variables.

Notes:

1. The set T is called the **index set** for the stochastic process. The only set T that need be considered for this appendix is $T = \{0, 1, 2, 3, \dots\}$, so the stochastic process can be described as the sequence of random variables $\{X_0, X_1, X_2, \dots\}$. When $T = \{0, 1, 2, 3, \dots\}$, the index is often identified with time and the stochastic process is called a discrete-time stochastic process. The random variable X_k is understood to be the stochastic process at time k .
2. It is assumed that the random variables in a stochastic process have a common range. This range is called the **state space** for the stochastic process. The state spaces in Chapter 10 are all finite, so the random variables X_k are all discrete random variables. If $X_k = i$, we will say that i is the **state** of the process at time k , or that the process is in state i at time (or step) k .
3. Notice that a stochastic process can be used to model movement between the states in the state space. For some element ω in the sample space S , the sequence $\{X_0(\omega), X_1(\omega), \dots\}$ will be a sequence of states in the state space—a sequence that will potentially be different for each element in S . Usually the dependence on the sample space is ignored and the stochastic process is treated as a sequence of states, and the process is said to move (or transition) between those states as time proceeds.
4. Since a stochastic process is a sequence of random variables, the actual state that the process occupies at any given time cannot be known. The goal therefore is to find the probability that the process is in a particular state at a particular time. This amounts to finding the probability mass function of each random variable X_k in the sequence that is the stochastic process.
5. When a discrete-time stochastic process has a finite state space, the probability mass function of each random variable X_k can be expressed as a probability vector \mathbf{x}_k . These probability vectors were used to define a Markov chain in Section 4.9.

In order for a discrete-time stochastic process $\{X_0, X_1, X_2, \dots\}$ to be a Markov chain, the state of the process at time $n + 1$ can depend only on the state of the process at time n . This is in contrast with a more general stochastic process, whose state at time n could depend on the entire history of the process. In terms of conditional probability, this property is

$$P(X_{n+1} = i | X_0 = j_0, X_1 = j_1, \dots, X_n = j) = P(X_{n+1} = i | X_n = j)$$

The probability on the right side of this equation is called the transition probability from state j to state i . In general, this transition probability can change depending on the time n . This is not the case for Markov chains considered in Chapter 10: the transition probabilities do not change with time, so the transition probability from state j to state i is

$$P(X_{n+1} = i | X_n = j) = p_{ij}$$

A Markov chain with constant transition probabilities is called a time-homogeneous Markov chain. Thus its definition is as follows.

DEFINITION

A **time-homogeneous Markov chain** is a discrete-time stochastic process whose transition probabilities satisfy

$$P(X_{n+1} = i | X_0 = j_0, X_1 = j_1, \dots, X_n = j) = P(X_{n+1} = i | X_n = j) = p_{ij}$$

for all times n and for all states i and j .

Using this definition, it is clear that, if the number of states is finite, then a transition matrix can be constructed that has the properties assumed in Section 10.1.

Proofs of Theorems

Mean Return Times

Theorem 3 in Section 10.3 connected the steady-state vector for a Markov chain with the mean return time to a state of the chain. Here is a statement of this theorem and a proof that relies on the law of total probability for expected value.

THEOREM 3

Let $X_n, n = 1, 2, \dots$ be an irreducible Markov chain with finite state space S . Let n_{ij} be the number of steps until the chain first visits state i given that the chain starts in state j , and let $t_{ii} = E[n_{ii}]$. Then

$$t_{ii} = \frac{1}{q_i}$$

where q_i is the entry in the steady-state vector \mathbf{q} corresponding to state i .

PROOF To find an expression for t_{ii} , first produce an equation involving t_{ij} by considering the first step of the chain X_1 . There are two possibilities: either $X_1 = i$ or $X_1 = k \neq i$. If $X_1 = i$, then it took exactly one step to visit state i and

$$E[n_{ij} | X_1 = i] = 1$$

If $X_1 = k \neq i$, the chain will take one step to reach state k , and then the expected number of steps the chain will make to first visit state i will be $E[n_{ik}] = t_{ik}$. Thus

$$E[n_{ij} | X_1 = k \neq i] = 1 + t_{ik}$$

By the law of total probability for expected value,

$$\begin{aligned} t_{ij} &= E[n_{ij}] \\ &= \sum_{k \in S} E[n_{ij} | X_1 = k] P(X_1 = k) \\ &= E[n_{ij} | X_1 = i] P(X_1 = i) + \sum_{k \neq i} E[n_{ij} | X_1 = k] P(X_1 = k) \\ &= 1 \cdot p_{ij} + \sum_{k \neq i} (1 + t_{ik}) p_{kj} \end{aligned}$$

$$\begin{aligned}
&= p_{ij} + \sum_{k \neq i} p_{kj} + \sum_{k \neq i} t_{ik} p_{kj} \\
&= 1 + \sum_{k \neq i} t_{ik} p_{kj} \\
&= 1 + \sum_{k \in S} t_{ik} p_{kj} - t_{ii} p_{ij}
\end{aligned}$$

Let T be the matrix whose (i, j) -element is t_{ij} , and let D be the diagonal matrix whose diagonal entries are t_{ii} . Then the final equality above may be written as

$$T_{ij} = 1 + (TP)_{ij} - (DP)_{ij} \quad (1)$$

If U is an appropriately sized matrix of ones, Equation (1) can be written in matrix form as

$$T = U + TP - DP = U + (T - D)P \quad (2)$$

Multiplying each side of Equation (2) by the steady-state vector \mathbf{q} and recalling that $P\mathbf{q} = \mathbf{q}$ gives

$$T\mathbf{q} = U\mathbf{q} + (T - D)P\mathbf{q} = U\mathbf{q} + (T - D)\mathbf{q} = U\mathbf{q} + T\mathbf{q} - D\mathbf{q}$$

so

$$U\mathbf{q} = D\mathbf{q} \quad (3)$$

Consider the entries in each of the vectors in Equation (3). Since U is a matrix of all 1's,

$$U\mathbf{q} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^n q_k \\ \sum_{k=1}^n q_k \\ \vdots \\ \sum_{k=1}^n q_k \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

since \mathbf{q} is a probability vector. Likewise,

$$D\mathbf{q} = \begin{bmatrix} t_{11} & 0 & \cdots & 0 \\ 0 & t_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & t_{nn} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = \begin{bmatrix} t_{11}q_1 \\ t_{22}q_2 \\ \vdots \\ t_{nn}q_n \end{bmatrix}$$

Equating corresponding entries in $U\mathbf{q}$ and $D\mathbf{q}$ gives $t_{ii}q_i = 1$, or

$$t_{ii} = \frac{1}{q_i}$$

Periodicity as a Class Property

In Section 10.4 it was stated that if two states belong to the same communication class, then their periods must be equal. A proof of this result follows.

THEOREM

Let i and j be two states of a Markov chain that are in the same communication class. Then the periods of i and j are equal.

PROOF Suppose that i and j are in the same communication class for the Markov chain X , that state i has period d_i , and that state j has period d_j . To simplify the exposition of the proof, the notation $(a^r)_{ij}$ will be used to refer to the (i, j) -entry in the matrix A^r . Since i and j are in the same communication class, there exist positive integers m and n such that $(p^m)_{ji} > 0$ and $(p^n)_{ij} > 0$. Let k be a positive integer such that $(p^k)_{jj} > 0$. In fact, $(p^{lk})_{jj} > 0$ for all integers $l > 1$. Now $(p^{n+lk+m})_{ii} > (p^n)_{ij}(p^{lk})_{jj}(p^m)_{ji} > 0$ for all integers $l > 1$, since a loop from state i to state i in $n + lk + m$ steps may be created in many ways, but one way is to proceed from state i to state j in n steps, then to loop from state j to state j l times using a loop of k steps each time, and then to return to state i in m steps. Since d_i is the period of state i , d_i must divide $n + lk + m$ for all integers $l > 1$. So d_i divides $n + k + m$ and $n + 2k + m$, and so divides $(n + 2k + m) - (n + k + m) = k$. Thus d_i is a common divisor of the set of all time steps k such that $(p^k)_{jj} > 0$. Since d_j is the *greatest* common divisor of the set of all time steps k such that $(p^k)_{jj} > 0$, $d_i \leq d_j$. A similar argument shows that $d_i \geq d_j$, so $d_i = d_j$. ■

The Fundamental Matrix

In Section 10.5, the number of visits v_{ij} to a transient state i that a Markov chain makes starting at the transient state j was studied. Specifically, the expected value $E[v_{ij}]$ was computed, and the fundamental matrix was defined as the matrix whose (i, j) -element is $m_{ij} = E[v_{ij}]$. The following theorem restates Theorem 6 in Section 10.5 in an equivalent form and provides a proof that relies on the law of total probability for expected value.

THEOREM 6

Let j and i be transient states of a Markov chain, and let Q be that portion of the transition matrix which governs movement between transient states. Let v_{ij} be the number of visits that the chain will make to state i given that the chain starts in state j , and let $m_{ij} = E[v_{ij}]$. Then the matrix M whose (i, j) -element is m_{ij} satisfies the equation

$$M = (I - Q)^{-1}$$

PROOF We produce an equation involving m_{ij} by conditioning on the first step of the chain X_1 . We consider two cases: $i \neq j$ and $i = j$. First assume that $i \neq j$ and suppose that $X_1 = k$. Then we see that

$$E[v_{ij} | X_1 = k] = E[v_{ik}] \quad (4)$$

if $i \neq j$. Now assume that $i = j$. Then the previous analysis is valid, but we must add one visit to i since the chain was at state i at time 0. Thus

$$E[v_{ii} | X_1 = k] = 1 + E[v_{ik}] \quad (5)$$

We may combine Equations (4) and (5) by introducing the following symbol, called the *Kronecker delta*:

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Notice that δ_{ij} is the (i, j) -element in the identity matrix I . We can write Equations (4) and (5) as

$$E[v_{ij} | X_1 = k] = \delta_{ij} + E[v_{ik}]$$

Thus, by the law of total probability for expected value,

$$\begin{aligned}
 m_{ij} &= E[v_{ij}] \\
 &= \sum_{k \in S} E[v_{ij} | X_1 = k] P(X_1 = k) \\
 &= \sum_{k \in S} (\delta_{ij} + E[v_{ik}]) P(X_1 = k) \\
 &= \delta_{ij} \sum_{k \in S} P(X_1 = k) + \sum_{k \in S} E[v_{ik}] P(X_1 = k) \\
 &= \delta_{ij} + \sum_{k \in S} E[v_{ik}] P(X_1 = k)
 \end{aligned}$$

Now note that if k is a recurrent state, then $E[v_{ik}] = 0$. Thus we only need to sum over transient states of the chain:

$$\begin{aligned}
 m_{ij} &= \delta_{ij} + \sum_{k \text{ transient}} E[v_{ik}] P(X_1 = k) \\
 &= \delta_{ij} + \sum_{k \text{ transient}} m_{ik} q_{kj}
 \end{aligned}$$

since j and k are transient states and Q is defined in the statement of the theorem. We may write the final equality above as

$$m_{ij} = I_{ij} + (MQ)_{ij}$$

or in matrix form as

$$M = I + MQ \quad (6)$$

We may rewrite Equation (6) as

$$M - MQ = M(I - Q) = I$$

so $(I - Q)$ is invertible by the Invertible Matrix Theorem, and $M = (I - Q)^{-1}$. ■

Absorption Probabilities

In Section 10.5, the probability that the chain was absorbed into a particular absorbing state was studied. The Markov chain was assumed to have only transient and absorbing states, j is a transient state, and i is an absorbing state of the chain. The probability a_{ij} that the chain is absorbed at state i given that the chain starts at state j was calculated, and it was shown that the matrix A whose (i, j) -element is a_{ij} satisfies $A = SM$, where M is the fundamental matrix and S is that portion of the transition matrix that governs movement from transient states to absorbing states. The following theorem restates this result, which was presented as Theorem 7 in Section 10.5. An alternative proof of this result is given that relies on the law of total probability.

THEOREM 7

Consider a Markov chain with finite state space whose states are either absorbing or transient. Suppose that j is a transient state and that i is an absorbing state of the chain, and let a_{ij} be the probability that the chain is absorbed at state i given that the chain starts in state j . Let A be the matrix whose (i, j) -element is a_{ij} . Then $A = SM$, where S and M are as defined above.

PROOF We again consider the first step of the chain X_1 . Let $X_1 = k$. There are three possibilities: k could be a transient state, k could be i , and k could be an absorbing state unequal to i . If k is transient, then

$$P(\text{absorption at } i | X_1 = k) = a_{ik}$$

If $k = i$, then

$$P(\text{absorption at } i | X_1 = k) = 1$$

while if k is an absorbing state other than i ,

$$P(\text{absorption at } i | X_1 = k) = 0$$

By the law of total probability,

$$\begin{aligned} a_{ij} &= P(\text{absorption at } i) \\ &= \sum_k P(\text{absorption at } i | X_1 = k) P(X_1 = k) \\ &= 1 \cdot P(X_1 = i) + \sum_{k \text{ transient}} P(\text{absorption at } i | X_1 = k) P(X_1 = k) \\ &= p_{ij} + \sum_{k \text{ transient}} a_{ik} p_{kj} \end{aligned}$$

Since j is transient and i is absorbing, $p_{ij} = s_{ij}$. Since in the final sum j and k are both transient, $p_{kj} = q_{kj}$. Thus the final equality may be written as

$$\begin{aligned} a_{ij} &= s_{ij} + \sum_{k \text{ transient}} a_{ik} q_{kj} \\ &= s_{ij} + (AQ)_{ij} \end{aligned}$$

or, in matrix form, as

$$A = S + AQ$$

This equation may be solved for A to find that $A = SM$. ■