Chapter 2: Markov Chains and Queues in Discrete Time

L. Breuer
University of Kent

1 Definition

Let X_n with $n \in \mathbb{N}_0$ denote random variables on a discrete space E. The sequence $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ is called a **stochastic chain**. If $\mathbb{P}$ is a probability measure $\mathcal{X}$ such that

$$\mathbb{P}(X_{n+1} = j | X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = j | X_n = i_n) \qquad (1)$$

for all $i_0, \ldots, i_n, j \in E$ and $n \in \mathbb{N}_0$, then the sequence $\mathcal{X}$ shall be called a **Markov chain** on E. The probability measure $\mathbb{P}$ is called the distribution of $\mathcal{X}$, and E is called the **state space** of $\mathcal{X}$.

If the conditional probabilities $\mathbb{P}(X_{n+1} = j | X_n = i_n)$ are independent of the time index $n \in \mathbb{N}_0$, then we call the Markov chain $\mathcal{X}$ **homogeneous** and denote

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

for all $i, j \in E$. The probability p_{ij} is called **transition probability** from state i to state j. The matrix $P := (p_{ij})_{i,j \in E}$ shall be called **transition matrix** of the chain $\mathcal{X}$. Condition (1) is referred to as the **Markov property**.

Example 1 If $(X_n : n \in \mathbb{N}_0)$ are random variables on a discrete space E, which are stochastically independent and identically distributed (shortly: iid), then the chain $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ is a homogeneous Markov chain.

Example 2 Discrete Random Walk
Set $E := \mathbb{Z}$ and let $(S_n : n \in \mathbb{N})$ be a sequence of iid random variables with values in $\mathbb{Z}$ and distribution π. Define $X_0 := 0$ and $X_n := \sum_{k=1}^{n} S_k$ for all $n \in \mathbb{N}$. Then the chain $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ is a homogeneous Markov chain with transition probabilities $p_{ij} = \pi_{j-i}$. This chain is called **discrete random walk**.

Example 3 Bernoulli process
Set $E := \mathbb{N}_0$ and choose any parameter $0 < p < 1$. The definitions $X_0 := 0$ as well as

$$p_{ij} := \begin{cases} p, & j = i + 1 \\ 1 - p, & j = i \end{cases}$$

for $i \in \mathbb{N}_0$ determine a homogeneous Markov chain $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$. It is called **Bernoulli process** with parameter p.

So far, al examples have been chosen as to be homogeneous. The following theorem shows that there is a good reason for this:

Theorem 1 *Be $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ a Markov chain on a discrete state space E. Then there is a homogeneous Markov chain $\mathcal{X}' = (X'_n : n \in \mathbb{N}_0)$ on the state space $E \times \mathbb{N}_0$ such that $X_n = pr_1(X'_n)$ for all $n \in \mathbb{N}_0$, with pr_1 denoting the projection to the first dimension.*

Proof: Let $\mathcal{X}$ be a Markov chain with transition probabilities

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

which may depend on the time instant n. Define the two–dimensional random variables $X'_n := (X_n, n)$ for all $n \in \mathbb{N}_0$ and denote the resulting distribution of the chain $\mathcal{X}' = (X'_n : n \in \mathbb{N}_0)$ by $\mathbb{P}'$. By definition we obtain $X_n = pr_1(X'_n)$ for all $n \in \mathbb{N}_0$.

Further $\mathbb{P}'(X'_0 = (i,k)) = \delta_{k0} \cdot \mathbb{P}(X_0 = i)$ holds for all $i \in E$, and all transition probabilities

$$p'_{(i,k),(j,l)} = \mathbb{P}'(X'_{k+1} = (j,l) | X'_k = (i,k)) = \delta_{l,k+1} \cdot p_{k;ij}$$

can be expressed without a time index. Hence the Markov chain $\mathcal{X}'$ is homogeneous.
□

Because of this result, we will from now on treat only homogeneous Markov chains and omit the adjective "homogeneous".

Let P denote the transition matrix of a Markov chain on E. Then as an immediate consequence of its definition we obtain $p_{ij} \in [0,1]$ for all $i, j \in E$ and $\sum_{j \in E} p_{ij} = 1$ for all $i \in E$. A matrix P with these properties is called a **stochastic matrix** on E. In the following we shall demonstrate that, given an initial distribution, a Markov chain is uniquely determined by its transition matrix. Thus any stochastic matrix defines a family of Markov chains.

Theorem 2 *Let $\mathcal{X}$ denote a homogeneous Markov chain on E with transition matrix P. Then the relation*

$$\mathbb{P}\left(X_{n+1}=j_1,\ldots,X_{n+m}=j_m|X_n=i\right)=p_{i,j_1}\cdot\ldots\cdot p_{j_{m-1},j_m}$$

holds for all $n\in\mathbb{N}_0$, $m\in\mathbb{N}$, and $i,j_1,\ldots,j_m\in E$.

Proof: This is easily shown by induction on m. For $m=1$ the statement holds by definition of P. For $m>1$ we can write

$$\begin{aligned}
\mathbb{P}(X_{n+1}=&j_1,\ldots,X_{n+m}=j_m|X_n=i)\\
&=\frac{\mathbb{P}\left(X_{n+1}=j_1,\ldots,X_{n+m}=j_m,X_n=i\right)}{\mathbb{P}\left(X_n=i\right)}\\
&=\frac{\mathbb{P}\left(X_{n+1}=j_1,\ldots,X_{n+m}=j_m,X_n=i\right)}{\mathbb{P}\left(X_{n+1}=j_1,\ldots,X_{n+m-1}=j_{m-1},X_n=i\right)}\\
&\qquad\times\frac{\mathbb{P}\left(X_{n+1}=j_1,\ldots,X_{n+m-1}=j_{m-1},X_n=i\right)}{\mathbb{P}\left(X_n=i\right)}\\
&=\mathbb{P}\left(X_{n+m}=j_m|X_n=i,X_{n+1}=j_1,\ldots,X_{n+m-1}=j_{m-1}\right)\\
&\qquad\times p_{i,j_1}\cdot\ldots\cdot p_{j_{m-2},j_{m-1}}\\
&=p_{j_{m-1},j_m}\cdot p_{i,j_1}\cdot\ldots\cdot p_{j_{m-2},j_{m-1}}
\end{aligned}$$

because of the induction hypothesis and the Markov property.
□

Let π be a probability distribution on E with $\mathbb{P}(X_0=i)=\pi_i$ for all $i\in E$. Then theorem 2 immediately yields

$$\mathbb{P}\left(X_0=j_0,X_1=j_1,\ldots,X_m=j_m\right)=\pi_{j_0}\cdot p_{j_0,j_1}\ldots p_{j_{m-1},j_m} \tag{2}$$

for all $m\in\mathbb{N}$ and $j_0,\ldots,j_m\in E$. The chain with this distribution $\mathbb{P}$ is denoted by $\mathcal{X}^\pi$ and called the π–**version** of $\mathcal{X}$. The probability measure π is called **initial distribution** for $\mathcal{X}$.

Theorem 2 and the extension theorem by Tulcea (see appendix 6) show that a Markov chain is uniquely determined by its transition matrix and its initial distribution. Whenever the initial distribution π is not important or understood from the context, we will simply write $\mathcal{X}$ instead of $\mathcal{X}^\pi$. However, in an exact manner the notation $\mathcal{X}$ denotes the family of all the versions $\mathcal{X}^\pi$ of $\mathcal{X}$, indexed by their initial distribution π.

Theorem 3 *Let $\mathcal{X}$ denote a homogeneous Markov chain with transition matrix P. Then the relation*

$$\mathbb{P}(X_{n+m} = j | X_n = i) = P^m(i, j)$$

holds for all $m, n \in \mathbb{N}_0$ and $i, j \in E$, with $P^m(i, j)$ denoting the (i, j)th entry of the mth power of the matrix P. In particular, P^0 equals the identity matrix.

Proof: This follows by induction on m. For $m = 1$ the statement holds by definition of P. For $m > 1$ we can write

$$\begin{aligned}
\mathbb{P}(X_{n+m} = j | X_n = i) &= \frac{\mathbb{P}\left(X_{n+m} = j, X_n = i\right)}{\mathbb{P}\left(X_n = i\right)} \\
&= \sum_{k \in E} \frac{\mathbb{P}\left(X_{n+m} = j, X_{n+m-1} = k, X_n = i\right)}{\mathbb{P}\left(X_{n+m-1} = k, X_n = i\right)} \\
&\quad \times \frac{\mathbb{P}\left(X_{n+m-1} = k, X_n = i\right)}{\mathbb{P}\left(X_n = i\right)} \\
&= \sum_{k \in E} \mathbb{P}\left(X_{n+m} = j | X_{n+m-1} = k, X_n = i\right) \cdot P^{m-1}(i, k) \\
&= \sum_{k \in E} p_{kj} \cdot P^{m-1}(i, k) = P^m(i, j)
\end{aligned}$$

because of the induction hypothesis and the Markov property.
□

Thus the probabilities for transitions in m steps are given by the mth power of the transition matrix P. The rule $P^{m+n} = P^m P^n$ for the multiplication of matrices and theorem 3 lead to the decompositions

$$\mathbb{P}(X_{m+n} = j | X_0 = i) = \sum_{k \in E} \mathbb{P}(X_m = k | X_0 = i) \cdot \mathbb{P}(X_n = j | X_0 = k)$$

which are known as the **Chapman–Kolmogorov equations**.

For later purposes we will need a relation closely related to the Markov property, which is called the **strong Markov property**. Let τ denote a random variable with values in $\mathbb{N}_0 \cup \{\infty\}$, such that the condition

$$\mathbb{P}(\tau \leq n | \mathcal{X}) = \mathbb{P}(\tau \leq n | X_0, \ldots, X_n) \tag{3}$$

holds for all $n \in \mathbb{N}_0$. Such a random variable is called a (discrete) **stopping time** for $\mathcal{X}$. The defining condition means that the probability for the event $\{\tau \leq$

$n\}$ depends only on the evolution of the chain until time n. In other words, the determination of a stopping time does not require any knowledge of the future. Now the strong Markov property is stated in

Theorem 4 *Let $\mathcal{X}$ denote a Markov chain and τ a stopping time for $\mathcal{X}$ with $\mathbb{P}(\tau < \infty) = 1$. Then the relation*

$$\mathbb{P}(X_{\tau+m} = j | X_0 = i_0, \ldots, X_\tau = i_\tau) = \mathbb{P}(X_m = j | X_0 = i_\tau)$$

holds for all $m \in \mathbb{N}$ and $i_0, \ldots, i_\tau, j \in E$.

Proof: The fact that the stopping time τ is finite and may assume only countably many values can be exploited in the transformation

$$\begin{aligned}
&\mathbb{P}(X_{\tau+m} = j | X_0 = i_0, \ldots, X_\tau = i_\tau) \\
&\qquad = \sum_{n=0}^{\infty} \mathbb{P}(\tau = n, X_{\tau+m} = j | X_0 = i_0, \ldots, X_\tau = i_\tau) \\
&\qquad = \sum_{n=0}^{\infty} \mathbb{P}(X_{\tau+m} = j | \tau = n, X_0 = i_0, \ldots, X_\tau = i_\tau) \\
&\qquad\qquad \times \mathbb{P}(\tau = n | X_0 = i_0, \ldots, X_\tau = i_\tau) \\
&\qquad = \sum_{n=0}^{\infty} \mathbb{P}(X_{n+m} = j | X_n = i_\tau) \cdot \mathbb{P}(\tau = n | \mathcal{X}) \\
&\qquad = \sum_{n=0}^{\infty} \mathbb{P}(\tau = n | \mathcal{X}) \cdot \mathbb{P}(X_m = j | X_0 = i_\tau)
\end{aligned}$$

which yields the statement, as τ is finite with probability one.
□

2 Classification of States

Let $\mathcal{X}$ denote a Markov chain with state space E and transition matrix P. We call a state $j \in E$ **accessible** from a state $i \in E$ if there is a number $m \in \mathbb{N}_0$ with $P(X_m = j | X_0 = i) > 0$. This relation shall be denoted by $i \to j$. If for two states $i, j \in E$, the relations $i \to j$ and $j \to i$ hold, then i and j are said to **communicate**, in notation $i \leftrightarrow j$.

Theorem 5 *The relation $\leftrightarrow$ of communication between states is an equivalence relation.*

Proof: Because of $P^0 = I$, communication is reflexive. Symmetry holds by definition. Thus it remains to show transitivity. For this, assume $i \leftrightarrow j$ and $j \leftrightarrow k$ for three states $i, j, k \in E$. This means that there are numbers $m, n \in \mathbb{N}_0$ with $P^m(i, j) > 0$ and $P^n(j, k) > 0$. Hence, by the Chapman–Kolmogorov equation, we obtain

$$\begin{aligned}\mathbb{P}(X_{m+n} = k|X_0 = i) &= \sum_{h\in E} \mathbb{P}(X_m = h|X_0 = i) \cdot \mathbb{P}(X_n = k|X_0 = h)\\ &\geq \mathbb{P}(X_m = j|X_0 = i) \cdot \mathbb{P}(X_n = k|X_0 = j) > 0\end{aligned}$$

which proves $i \to k$. The remaining proof of $k \to i$ is completely analogous.
$\square$

Because of this result and the countability, we can divide the state space E of a Markov chain into a partition of countably many equivalence classes with respect to the communication of states. Any such equivalence class shall be called **communication class**. A communication class $C \subset E$ that does not allow access to states outside itself, i.e. for which the implication

$$i \to j, \quad i \in C \qquad \Rightarrow \qquad j \in C$$

holds, is called **closed**. If a closed equivalence class consists only of one state, then this state shall be called **absorbing**. If a Markov chain has only one communication class, i.e. if all states are communicating, then it is called **irreducible**. Otherwise it is called **reducible**.

Example 4 Let $\mathcal{X}$ denote a discrete random walk (see example 2) with the specification $\pi_1 = p$ and $\pi_{-1} = 1 - p$ for some parameter $0 < p < 1$. Then $\mathcal{X}$ is irreducible.

Example 5 The Bernoulli process (see example 3) with non–trivial parameter $0 < p < 1$ is to the highest degree reducible. Every state $x \in \mathbb{N}_0$ forms an own communication class. None of these is closed, thus there are no absorbing states.

Theorem 6 *Be $\mathcal{X}$ a Markov chain with state space E and transition matrix P. Let $C = \{c_n : n \in I\} \subset E$ with $I \subset \mathbb{N}$ be a closed communication class. Define the matrix P' by its entries $p'_{ij} := p_{c_i,c_j}$ for all $i, j \in I$. Then P' is stochastic.*

Proof: By definition, $p'_{ij} \in [0,1]$ for all $i,j \in I$. Since C is closed, $p_{c_i,k} = 0$ for all $i \in I$ and $k \notin C$. This implies

$$\sum_{j\in I} p'_{ij} = \sum_{j\in I} p_{c_i,c_j} = 1 - \sum_{k\notin C} p_{c_i,k} = 1$$

for all $i \in I$, as P is stochastic.
□

Thus the restriction of a Markov chain $\mathcal{X}$ with state space E to the states of one of its closed communication classes C defines a new Markov chain with state space C. If the states are relabeled according to their affiliation to a communication class, the transition matrix of $\mathcal{X}$ can be displayed in a block matrix form as

$$P = \begin{bmatrix} Q & Q_1 & Q_2 & Q_3 & Q_4 & \dots \\ 0 & P_1 & 0 & 0 & 0 & \dots \\ 0 & 0 & P_2 & 0 & 0 & \dots \\ 0 & 0 & 0 & P_3 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \end{bmatrix} \tag{4}$$

with P_n being stochastic matrices on the closed communication classes C_n. The first row contains the transition probabilities starting from communication classes that are not closed.

Let $\mathcal{X}$ denote a Markov chain with state space E. In the rest of this section we shall investigate distribution and expectation of the following random variables: Define τ_j as the stopping time of the **first visit** to the state $j \in E$, i.e.

$$\tau_j := \min\{n \in \mathbb{N} : X_n = j\}$$

Denote the distribution of τ_j by

$$F_k(i,j) := \mathbb{P}(\tau_j = k | X_0 = i)$$

for all $i,j \in E$ and $k \in \mathbb{N}$.

Lemma 1 The conditional distribution of the first visit to the state $j \in E$, given an initial state $X_0 = i$, can be determined iteratively by

$$F_k(i,j) = \begin{cases} p_{ij}, & k = 1 \\ \sum_{h\neq j} p_{ih} F_{k-1}(h,j), & k \geq 2 \end{cases}$$

for all $i,j \in E$.

Proof: For $k = 1$, the definition yields

$$F_1(i,j) = \mathbb{P}(\tau_j = 1|X_0 = i) = \mathbb{P}(X_1 = j|X_0 = i) = p_{ij}$$

for all $i, j \in E$. For $k \geq 2$, conditioning upon X_1 yields

$$F_k(i,j) = \mathbb{P}(X_1 \neq j, \ldots, X_{k-1} \neq j, X_k = j|X_0 = i)$$

$$\begin{aligned} &= \sum_{h \neq j} \mathbb{P}(X_1 = h|X_0 = i) \\ &\quad \times \mathbb{P}(X_2 \neq j, \ldots, X_{k-1} \neq j, X_k = j|X_0 = i, X_1 = h) \\ &= \sum_{h \neq j} p_{ih} \cdot \mathbb{P}(X_1 \neq j, \ldots, X_{k-2} \neq j, X_{k-1} = j|X_0 = h) \end{aligned}$$

due to the Markov property.
□

Now define

$$f_{ij} := \mathbb{P}(\tau_j < \infty|X_0 = i) = \sum_{k=1}^{\infty} F_k(i,j) \tag{5}$$

for all $i, j \in E$, which represents the probability of ever visiting state j after beginning in state i. Summing up over all $k \in \mathbb{N}$ in the formula of Lemma 1 leads to

$$f_{ij} = p_{ij} + \sum_{h \neq j} p_{ih} f_{hj} \tag{6}$$

for all $i, j \in E$. The proof is left as an exercise.

Define N_j as the random variable of the **total number of visits** to the state $j \in E$. Expression (6) is useful for computing the distribution of N_j:

Theorem 7 *Let $\mathcal{X}$ denote a Markov chain with state space E. The total number of visits to a state $j \in E$ under the condition that the chain starts in state i is given by*

$$\mathbb{P}(N_j = m|X_0 = j) = f_{jj}^{m-1}(1 - f_{jj})$$

and for $i \neq j$

$$\mathbb{P}(N_j = m|X_0 = i) = \begin{cases} 1 - f_{ij}, & m = 0 \\ f_{ij} f_{jj}^{m-1}(1 - f_{jj}), & m \geq 1 \end{cases}$$

Thus the distribution of N_j is modified geometric.

Proof: Define $\tau_j^{(1)} := \tau_j$ and $\tau_j^{(k+1)} := \min\{n > \tau_j^{(k)} : X_n = j\}$ for all $k \in \mathbb{N}$, with the convention that $\min \emptyset = \infty$. Note that $\tau_j^{(k)} = \infty$ implies $\tau_j^{(l)} = \infty$ for all $l > k$.

Then the sequence $(\tau_j^{(k)} : k \in \mathbb{N})$ is a sequence of stopping times. The event $\{N_j = m\}$ is the same as the intersection of the events $\{\tau_j^{(k)} < \infty\}$ for $k = 1, \ldots, M$ and $\{\tau_j^{(M+1)} = \infty\}$, with $M = m$ if $i \neq j$ and $M = m - 1$ if $i = j$. Now this event can be further described by the intersection of the events $\{\tau_j^{(k+1)} - \tau_j^{(k)} < \infty\}$ for $k = 0, \ldots, M-1$ and $\{\tau_j^{(M+1)} - \tau_j^{(M)} = \infty\}$, with M as above and the convention $\tau_j^{(0)} := 0$.

The subevent $\{\tau_j^{(k+1)} - \tau_j^{(k)} < \infty\}$ has probability f_{ij} for $k = 0$ and because of the strong Markov property (see theorem 4) probability f_{jj} for $k > 0$. The probability for $\{\tau_j^{(M+1)} - \tau_j^{(M)} = \infty\}$ is $1 - f_{ij}$ for $M = 0$ and $1 - f_{jj}$ for $M > 0$. Once more the strong Markov property is the reason for independence of the subevents. Now multiplication of the probabilities leads to the formulae in the statement.

□

Summing over all m in the above theorem leads to

Corollary 1 For all $j \in E$, the zero–one law

$$\mathbb{P}(N_j < \infty | X_0 = j) = \begin{cases} 1, & f_{jj} < 1 \\ 0, & f_{jj} = 1 \end{cases}$$

holds, i.e. depending on f_{jj} there are almost certainly infinitely many visits to a state $j \in E$.

This result gives rise to the following definitions: A state $j \in E$ is called **recurrent** if $f_{jj} = 1$ and **transient** otherwise. Let us further define the **potential matrix** $R = (r_{ij})_{i,j \in E}$ of the Markov chain by its entries

$$r_{ij} := \mathbb{E}(N_j | X_0 = i)$$

for all $i, j \in E$. Thus an entry r_{ij} gives the expected number of visits to the state $j \in E$ under the condition that the chain starts in state $i \in E$. As such, r_{ij} can be computed by

$$r_{ij} = \sum_{n=0}^{\infty} P^n(i,j) \tag{7}$$

for all $i, j \in E$. The results in theorem 7 and corollary 1 yield

Corollary 2 For all $i, j \in E$ the relations

$$r_{jj} = (1 - f_{jj})^{-1} \qquad \text{and} \qquad r_{ij} = f_{ij} r_{jj}$$

hold, with the conventions $0^{-1} := \infty$ and $0 \cdot \infty := 0$ included. In particular, the expected number r_{jj} of visits to the state $j \in E$ is finite if j is transient and infinite if j is recurrent.

Theorem 8 *Recurrence and transience of states are class properties with respect to the relation $\leftrightarrow$. Furthermore, a recurrent communication class is always closed.*

Proof: Assume that $i \in E$ is transient and $i \leftrightarrow j$. Then there are numbers $m, n \in \mathbb{N}$ with $0 < P^m(i,j) \le 1$ and $0 < P^n(j,i) \le 1$. The inequalities

$$\sum_{k=0}^{\infty} P^k(i,i) \ge \sum_{h=0}^{\infty} P^{m+h+n}(i,i) \ge P^m(i,j) P^n(j,i) \sum_{k=0}^{\infty} P^k(j,j)$$

now imply $r_{jj} < \infty$ because of representation (7). According to corollary 2 this means that j is transient, too.

If j is recurrent, then the same inequalities lead to

$$r_{ii} \ge P^m(i,j) P^n(j,i) r_{jj} = \infty$$

which signifies that i is recurrent, too. Since the above arguments are symmetric in i and j, the proof of the first statement is complete.

For the second statement assume that $i \in E$ belongs to a communication class $C \subset E$ and $p_{ij} > 0$ for some state $j \in E \setminus C$. Then

$$f_{ii} = p_{ii} + \sum_{h \ne i} p_{ih} f_{hi} \le 1 - p_{ij} < 1$$

according to formula (6), since $f_{ji} = 0$ (otherwise $i \leftrightarrow j$). Thus i is transient, which proves the second statement.

□

Theorem 9 *If the state $j \in E$ is transient, then $\lim_{n\to\infty} P^n(i,j) = 0$, regardless of the initial state $i \in E$.*

Proof: If the state j is transient, then the first equation in corollary 2 yields $r_{jj} < \infty$. The second equation in the same corollary now implies $r_{ij} < \infty$, which by the representation (7) completes the proof.

□

3 Stationary Distributions

Let $\mathcal{X}$ denote a Markov chain with state space E and π a measure on E. If $\mathbb{P}(X_n = i) = \mathbb{P}(X_0 = i) = \pi_i$ for all $n \in \mathbb{N}$ and $i \in E$, then $\mathcal{X}^\pi$ is called **stationary**, and π is called a **stationary measure** for $\mathcal{X}$. If furthermore π is a probability measure, then it is called **stationary distribution** for $\mathcal{X}$.

Theorem 10 *Let $\mathcal{X}$ denote a Markov chain with state space E and transition matrix P. Further, let π denote a probability distribution on E with $\pi P = \pi$, i.e.*

$$\pi_i = \sum_{j \in E} \pi_j p_{ji} \qquad \text{and} \qquad \sum_{j \in E} \pi_j = 1$$

for all $i \in E$. Then π is a stationary distribution for X. If π is a stationary distribution for $\mathcal{X}$, then $\pi P = \pi$ holds.

Proof: Let $\mathbb{P}(X_0 = i) = \pi_i$ for all $i \in E$. Then $\mathbb{P}(X_n = i) = \mathbb{P}(X_0 = i)$ for all $n \in \mathbb{N}$ and $i \in E$ follows by induction on n. The case $n = 1$ holds by assumption, and the induction step follows by induction hypothesis and the Markov property. The last statement is obvious.
□

The following examples show some features of stationary distributions:

Example 6 Let the transition matrix of a Markov chain $\mathcal{X}$ be given by

$$P = \begin{pmatrix} 0.8 & 0.2 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.6 & 0.4 \end{pmatrix}$$

Then $\pi = (0.5, 0.5, 0, 0)$, $\pi' = (0, 0, 0.5, 0.5)$ as well as any linear combination of them are stationary distributions for $\mathcal{X}$. This shows that a stationary distribution does not need to be unique.

Example 7 Bernoulli process (see example 1)
The transition matrix of a Bernoulli process has the structure

$$P = \begin{pmatrix} 1-p & p & 0 & 0 & \dots \\ 0 & 1-p & p & 0 & \ddots \\ 0 & 0 & 1-p & p & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

Hence $\pi P = \pi$ implies first

$$\pi_0 \cdot (1-p) = \pi_0 \quad \Rightarrow \quad \pi_0 = 0$$

since $0 < p < 1$. Assume that $\pi_n = 0$ for any $n \in \mathbb{N}_0$. This and the condition $\pi P = \pi$ further imply for π_{n+1}

$$\pi_n \cdot p + \pi_{n+1} \cdot (1-p) = \pi_{n+1} \quad \Rightarrow \quad \pi_{n+1} = 0$$

which completes an induction argument proving $\pi_n = 0$ for all $n \in \mathbb{N}_0$. Hence the Bernoulli process does not have a stationary distribution.

Example 8 The solution of $\pi P = \pi$ and $\sum_{j \in E} \pi_j = 1$ is unique for

$$P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$$

with $0 < p < 1$. Thus there are transition matrices which have exactly one stationary distribution.

The question of existence and uniqueness of a stationary distribution is one of the most important problems in the theory of Markov chains. A simple answer can be given in the transient case (cf. example 7):

Theorem 11 *A transient Markov chain (i.e. a Markov chain with transient states only) has no stationary distribution.*

Proof: Assume that $\pi P = \pi$ holds for some distribution π and take any enumeration $E = (s_n : n \in \mathbb{N})$ of the state space E. Choose any index $m \in \mathbb{N}$ with $\pi_{s_m} > 0$. Since $\sum_{n=1}^{\infty} \pi_{s_n} = 1$ is bounded, there is an index $M > m$ such that $\sum_{n=M}^{\infty} \pi_{s_n} < \pi_{s_m}$. Set $\varepsilon := \pi_{s_m} - \sum_{n=M}^{\infty} \pi_{s_n}$. According to theorem 9, there is an index $N \in \mathbb{N}$ such that $P^n(s_i, s_m) < \varepsilon$ for all $i \leq M$ and $n \geq N$. Then the stationarity of π implies

$$\begin{aligned} \pi_{s_m} &= \sum_{i=1}^{\infty} \pi_{s_i} P^N(s_i, s_m) = \sum_{i=1}^{M-1} \pi_{s_i} P^N(s_i, s_m) + \sum_{i=M}^{\infty} \pi_{s_i} P^N(s_i, s_m) \\ &< \varepsilon + \sum_{i=M}^{\infty} \pi_{s_i} = \pi_{s_m} \end{aligned}$$

which is a contradiction.

□

For the recurrent case, a finer distinction will be necessary. While the expected total number r_{jj} of visits to a recurrent state $j \in E$ is always infinite (see corollary 2), there are differences in the rate of visits to a recurrent state. In order to describe these, define $N_i(n)$ as the number of visits to state i until time n. Further define for a recurrent state $i \in E$ the mean time

$$m_i := \mathbb{E}(\tau_i | X_0 = i)$$

until the first visit to i (after time zero) under the condition that the chain starts in i. By definition $m_i > 0$ for all $i \in E$. The elementary renewal theorem (which will be proven later) states that

$$\lim_{n \to \infty} \frac{\mathbb{E}(N_i(n) | X_0 = j)}{n} = \frac{1}{m_i} \tag{8}$$

for all recurrent $i \in E$ and independently of $j \in E$ provided $j \leftrightarrow i$, with the convention of $1/\infty := 0$. Thus the asymptotic rate of visits to a recurrent state is determined by the mean recurrence time of this state. This gives reason to the following definition: A recurrent state $i \in E$ with $m_i = \mathbb{E}(\tau_i | X_0 = i) < \infty$ will be called **positive recurrent**, otherwise i is called **null recurrent**. The distinction between positive and null recurrence is supported by the equivalence relation $\leftrightarrow$, as shown in

Theorem 12 *Positive recurrence and null recurrence are class properties with respect to the relation of communication between states.*

Proof: Assume that $i \leftrightarrow j$ for two states $i, j \in E$ and i is null recurrent. Thus there are numbers $m, n \in \mathbb{N}$ with $P^n(i,j) > 0$ and $P^m(j,i) > 0$. Because of the representation $\mathbb{E}(N_i(k)|X_0 = i) = \sum_{l=0}^{k} P^l(i,i)$, we obtain

$$\begin{aligned}
0 &= \lim_{k\to\infty} \frac{\sum_{l=0}^{k} P^l(i,i)}{k} \\
&\geq \lim_{k\to\infty} \frac{\sum_{l=0}^{k-m-n} P^l(j,j)}{k} \cdot P^n(i,j)P^m(j,i) \\
&= \lim_{k\to\infty} \frac{k-m-n}{k} \cdot \frac{\sum_{l=0}^{k-m-n} P^l(j,j)}{k-m-n} \cdot P^n(i,j)P^m(j,i) \\
&= \lim_{k\to\infty} \frac{\sum_{l=0}^{k} P^l(j,j)}{k} \cdot P^n(i,j)P^m(j,i) \\
&= \frac{P^n(i,j)P^m(j,i)}{m_j}
\end{aligned}$$

and thus $m_j = \infty$, which signifies the null recurrence of j.
□

Thus we can call a communication class positive recurrent or null recurrent. In the former case, a construction of a stationary distribution is given in

Theorem 13 *Let $i \in E$ be positive recurrent and define the mean first visit time $m_i := \mathbb{E}(\tau_i|X_0 = i)$. Then a stationary distribution π is given by*

$$\pi_j := m_i^{-1} \cdot \sum_{n=0}^{\infty} \mathbb{P}(X_n = j, \tau_i > n|X_0 = i)$$

for all $j \in E$. In particular, $\pi_i = m_i^{-1}$ and $\pi_k = 0$ for all states k outside of the communication class belonging to i.

Proof: First of all, π is a probability measure since

$$\begin{aligned}
\sum_{j\in E}\sum_{n=0}^{\infty} \mathbb{P}(X_n = j, \tau_i > n|X_0 = i) &= \sum_{n=0}^{\infty}\sum_{j\in E} \mathbb{P}(X_n = j, \tau_i > n|X_0 = i) \\
&= \sum_{n=0}^{\infty} \mathbb{P}(\tau_i > n|X_0 = i) = m_i
\end{aligned}$$

The particular statements in the theorem are obvious from theorem 8 and the definition of π. The stationarity of π is shown as follows. First we obtain

$$\begin{aligned}\pi_j &= m_i^{-1} \cdot \sum_{n=0}^{\infty} \mathbb{P}(X_n = j, \tau_i > n | X_0 = i) \\ &= m_i^{-1} \cdot \sum_{n=1}^{\infty} \mathbb{P}(X_n = j, \tau_i \geq n | X_0 = i) \\ &= m_i^{-1} \cdot \sum_{n=1}^{\infty} \mathbb{P}(X_n = j, \tau_i > n-1 | X_0 = i)\end{aligned}$$

since $X_0 = X_{\tau_i} = i$ in the conditioning set $\{X_0 = i\}$. Because of

$$\begin{aligned}&\mathbb{P}(X_n = j, \tau_i > n-1 | X_0 = i) \\ &\quad = \frac{\mathbb{P}(X_n = j, \tau_i > n-1, X_0 = i)}{\mathbb{P}(X_0 = i)} \\ &\quad = \sum_{k \in E} \frac{\mathbb{P}(X_n = j, X_{n-1} = k, \tau_i > n-1, X_0 = i)}{\mathbb{P}(X_0 = i)} \\ &\quad = \sum_{k \in E} \frac{\mathbb{P}(X_n = j, X_{n-1} = k, \tau_i > n-1, X_0 = i)}{\mathbb{P}(X_{n-1} = k, \tau_i > n-1, X_0 = i)} \\ &\qquad \times \frac{\mathbb{P}(X_{n-1} = k, \tau_i > n-1, X_0 = i)}{\mathbb{P}(X_0 = i)} \\ &\quad = \sum_{k \in E} p_{kj} \mathbb{P}(X_{n-1} = k, \tau_i > n-1 | X_0 = i)\end{aligned}$$

we can transform further

$$\pi_j = m_i^{-1} \cdot \sum_{n=1}^{\infty} \sum_{k \in E} p_{kj} \mathbb{P}(X_{n-1} = k, \tau_i > n-1 | X_0 = i)$$

$$= \sum_{k \in E} p_{kj} \cdot m_i^{-1} \sum_{n=0}^{\infty} \mathbb{P}(X_n = k, \tau_i > n | X_0 = i) = \sum_{k \in E} \pi_k p_{kj}$$

which completes the proof.
□

Theorem 14 *Let $\mathcal{X}$ denote an irreducible, positive recurrent Markov chain. Then $\mathcal{X}$ has a unique stationary distribution.*

Proof: Existence has been shown in theorem 13. Uniqueness of the stationary distribution can be seen as follows. Let π denote the stationary distribution as constructed in theorem 13 and i the positive recurrent state that served as recurrence point for π. Further, let ν denote any stationary distribution for $\mathcal{X}$. Then there is a state $j \in E$ with $\nu_j > 0$ and a number $m \in \mathbb{N}$ with $P^m(j,i) > 0$, since $\mathcal{X}$ is irreducible. Consequently we obtain

$$\nu_i = \sum_{k \in E} \nu_k P^m(k,i) \geq \nu_j P^m(j,i) > 0$$

Hence we can multiply ν by a skalar factor c such that $c \cdot \nu_i = \pi_i = 1/m_i$. Denote $\tilde{\nu} := c \cdot \nu$.

Let $\tilde{P}$ denote the transition matrix P without the ith column, i.e. we define the (j,k)th entry of $\tilde{P}$ by $\tilde{p}_{jk} = p_{jk}$ if $k \neq i$ and zero otherwise. Denote further the Dirac measure on i by δ^i, i.e. $\delta^i_j = 1$ if $i = j$ and zero otherwise. Then the stationary distribution π can be represented by $\pi = m_i^{-1} \cdot \delta^i \sum_{n=0}^{\infty} \tilde{P}^n$.

We first claim that $m_i \tilde{\nu} = \delta^i + m_i \tilde{\nu} \tilde{P}$. This is clear for the entry $\tilde{\nu}_i$ and easily seen for $\tilde{\nu}_j$ with $j \neq i$ because in this case $(\tilde{\nu}\tilde{P})_j = c \cdot (\nu P)_j = \tilde{\nu}_j$. Now we can proceed with the same argument to see that

$$\begin{aligned} m_i \tilde{\nu} &= \delta^i + (\delta^i + m_i \tilde{\nu} \tilde{P}) \tilde{P} = \delta^i + \delta^i \tilde{P} + m_i \tilde{\nu} \tilde{P}^2 = \dots \\ &= \delta^i \sum_{n=0}^{\infty} \tilde{P}^n = m_i \pi \end{aligned}$$

Hence $\tilde{\nu}$ already is a probability measure and the skalar factor must be $c = 1$. This yields $\nu = \tilde{\nu} = \pi$ and thus the statement.
□

Remark 1 At a closer look the assumption of irreducibility may be relaxed to some extend. For example, if there is exactly one closed positive recurrent communication class and a set of transient and inaccessible states (i.e. states j for which there is no state i with $i \to j$), then the above statement still holds although $\mathcal{X}$ is not irreducible.

A first consequence of the uniqueness is the following simpler representation of the stationary distribution:

Theorem 15 *Let $\mathcal{X}$ denote an irreducible, positive recurrent Markov chain. Then the stationary distribution π of $\mathcal{X}$ is given by*

$$\pi_j = m_j^{-1} = \frac{1}{\mathbb{E}(\tau_j | X_0 = j)}$$

for all $j \in E$.

Proof: Since all states in E are positive recurrent, the construction in theorem 13 can be pursued for any inital state j. This yields $\pi_j = m_j^{-1}$ for all $j \in E$. The statement now follows from the uniqueness of the stationary distribution.
□

Corollary 3 For an irreducible, positive recurrent Markov chain, the stationary probability π_j of a state j coincides with its asymptotic rate of recurrence, i.e.

$$\lim_{n\to\infty} \frac{\mathbb{E}(N_j(n)|X_0 = i)}{n} = \pi_j$$

for all $j \in E$ and independently of $i \in E$. Further, if an asymptotic distribution $p = \lim_{n\to\infty} \mathbb{P}(X_n = .)$ does exist, then it coincides with the stationary distribution. In particular, it is independent of the initial distribution of $\mathcal{X}$.

Proof: The first statement immediately follows from equation (8). For the second statement, it suffices to employ $\mathbb{E}(N_j(n)|X_0 = i) = \sum_{l=0}^{n} P^l(i,j)$. If an asymptotic distribution p does exist, then for any initial distribution ν we obtain

$$\begin{aligned} p_j &= \lim_{n\to\infty} (\nu P^n)_j = \sum_{i\in E} \nu_i \lim_{n\to\infty} P^n(i,j) \\ &= \sum_{i\in E} \nu_i \lim_{n\to\infty} \frac{\sum_{l=0}^{n} P^l(i,j)}{n} = \sum_{i\in E} \nu_i \pi_j \\ &= \pi_j \end{aligned}$$

independently of ν.
□

4 Restricted Markov Chains

Now let $F \subset E$ denote any subset of the state space E. Define $\tau_F(k)$ to be the stopping time of the kth visit of $\mathcal{X}$ to the set F, i.e.

$$\tau_F(k+1) := \min\{n > \tau_F(k) : X_n \in F\}$$

with $\tau_F(0) := 0$. If $\mathcal{X}$ is recurrent, then the strong Markov property (theorem 4) ensures that the chain $\mathcal{X}^F = (X_n^F : n \in \mathbb{N})$ with $X_n^F := X_{\tau_F(n)}$ is a recurrent Markov chain, too. It is called the Markov chain restricted to F. In case of positive recurrence, we can obtain the stationary distribution of $\mathcal{X}^F$ from the stationary distribution of $\mathcal{X}$ in a simple manner:

Theorem 16 *If the Markov chain $\mathcal{X}$ is positive recurrent, then the stationary distribution of $\mathcal{X}^F$ is given by*

$$\pi_j^F = \frac{\pi_j}{\sum_{k\in F} \pi_k}$$

for all $j \in F$.

Proof: Choose any state $i \in F$ and recall from theorem 13 the expression

$$\pi_j := m_i^{-1} \cdot \sum_{n=0}^{\infty} \mathbb{P}(X_n = j, \tau_i > n | X_0 = i)$$

which holds for all $j \in F$. For π_j^F we can perform the same construction with respect to the chain $\mathcal{X}^F$. By the definition of $\mathcal{X}^F$ it is clear that the number of visits to the state j between two consecutive visits to i is the same for the chains $\mathcal{X}$ and $\mathcal{X}^F$. Hence the sum expression for π_j^F, which is the expectation of that number of visits, remains the same as for π_j. The other factor m_i^{-1} in the formula above is independent of j and serves only as a normalization constant, i.e. in order to secure that $\sum_{j\in E} \pi_j = 1$. Hence for a construction of π_j^F with respect to $\mathcal{X}^F$ this needs to be replaced by $(m_i \cdot \sum_{k\in F} \pi_k)^{-1}$, which then yields the statement.
□

Theorem 17 *Let $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ denote an irreducible and positive recurrent Markov chain with discrete state space E. Further let $F \subset E$ denote any subset of E, and $\mathcal{X}^F$ the Markov chain restricted to F. Denote*

$$\tau_F := \min\{n \in \mathbb{N} : X_n \in F\}$$

Then a measure ν on E is stationary for $\mathcal{X}$ if and only if $\nu' = (\nu_i : i \in F)$ is stationary for $\mathcal{X}^F$ and

$$\nu_j = \sum_{k\in F} \nu_k \sum_{n=0}^{\infty} \mathbb{P}(X_n = j, \tau_F > n | X_0 = k) \tag{9}$$

for all $j \in E \setminus F$.

Proof: Due to theorem 16 it suffices to prove equation (9) for $j \in E \setminus F$. Choose any state $i \in F$ and define

$$\tau_i := \min\{n \in \mathbb{N} : X_n = i\}$$

According to theorem 13 the stationary measure v for $\mathcal{X}$ is given by

$$\nu_j = \nu_i \cdot \sum_{n=0}^{\infty} \mathbb{P}(X_n = j, \tau_i > n | X_0 = i) = \nu_i \cdot \mathbb{E}_i \left(\sum_{n=0}^{\tau_i - 1} \mathbf{1}_{X_n = j} \right)$$

for $j \in E \setminus F$, where $\mathbb{E}_i$ denotes the conditional expectation given $X_0 = i$. Define further

$$\tau_i^F := \min\{n \in \mathbb{N} : X_n^F = i\}$$

Because of the strong Markov property we can proceed as

$$\begin{aligned} \nu_j &= \nu_i \cdot \mathbb{E}_i \left(\sum_{n=0}^{\tau_i^F - 1} \mathbb{E}_{X_n^F} \sum_{m=0}^{\tau_F - 1} \mathbf{1}_{X_m = j} \right) \\ &= \nu_i \cdot \sum_{k \in F} \mathbb{E}_i \left(\sum_{n=0}^{\tau_i^F - 1} \mathbf{1}_{X_n^F = k} \right) \cdot \mathbb{E}_k \left(\sum_{m=0}^{\tau_F - 1} \mathbf{1}_{X_m = j} \right) \end{aligned}$$

Regarding the restricted Markov chain $\mathcal{X}^F$, theorem 13 states that

$$\mathbb{E}_i \left(\sum_{n=0}^{\tau_i^F - 1} \mathbf{1}_{X_n^F = k} \right) = \sum_{n=0}^{\infty} \mathbb{P}(X_n^F = k, \tau_i^F > n | X_0^F = i) = \frac{\nu_k}{\nu_i}$$

for all $k \in F$. Hence we obtain

$$\nu_j = \sum_{k \in F} \nu_k \sum_{n=0}^{\infty} \mathbb{P}(X_n = j, \tau_F > n | X_0 = k)$$

which was to be proven.
□

5 Conditions for Positive Recurrence

In the third part of this course we will need some results on the behaviour of a Markov chain on a finite subset of its state space. As a first fundamental result we state

Theorem 18 *An irreducible Markov chain with finite state space F is positive recurrent.*

Proof: For all $n \in \mathbb{N}$ and $i \in F$ we have $\sum_{j \in E} P^n(i,j) = 1$. Hence it is not possible that $\lim_{n\to\infty} P^n(i,j) = 0$ for all $j \in F$. Thus there is one state $h \in F$ such that $r_{hh} = \sum_{n=0}^{\infty} P^n(h,h) = \infty$, which means by corollary 2 that h is recurrent and by irreducibility that the chain is recurrent.

If the chain were null recurrent, then according to the relation in (8)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P^k(i,j) = 0$$

would hold for all $j \in F$, independently of i because of irreducibility. But this would imply that $\lim_{n\to\infty} P^n(i,j) = 0$ for all $j \in F$, which contradicts our first observation in this proof. Hence the chain must be positive recurrent.
□

For irreducible Markov chains the condition $\mathbb{E}(\tau_i | X_0 = i) < \infty$ implies positive recurrence of state i and hence positive recurrence of the whole chain. Writing τ_F for the time of the first visit to the set F, we now can state the following generalization of this condition:

Theorem 19 *Let $\mathcal{X}$ denote an irreducible Markov chain with state space E and be $F \subset E$ a finite subset of E. The chain $\mathcal{X}$ is positive recurrent if and only if $\mathbb{E}(\tau_F | X_0 = i) < \infty$ for all $i \in F$.*

Proof: If $\mathcal{X}$ is positive recurrent, then $\mathbb{E}(\tau_F | X_0 = i) \le \mathbb{E}(\tau_i | X_0 = i) < \infty$ for all $i \in F$, by the definition of positive recurrence.

Now assume that $\mathbb{E}(\tau_F | X_0 = i) < \infty$ for all $i \in F$. Define the stopping times $\sigma(i) := \min\{k \in \mathbb{N} : X_k^F = i\}$ and random variables $Y_k := \tau_F(k) - \tau_F(k-1)$. Since F is finite, $m := \max_{j \in F} \mathbb{E}(\tau_F | X_0 = j) < \infty$. We shall denote the conditional expectation given $X_0 = i$ by $\mathbb{E}_i$. For $i \in F$ we now obtain

$$\begin{aligned} \mathbb{E}(\tau_i | X_0 = i) &= \mathbb{E}_i \left(\sum_{k=1}^{\sigma(i)} Y_k \right) = \sum_{k=1}^{\infty} \mathbb{E}_i \left(\mathbb{E}(Y_k | X_{\tau_F(k-1)}) \cdot 1_{k \le \sigma(i)} \right) \\ &\le m \cdot \sum_{k=1}^{\infty} \mathbb{P}(\sigma(i) \ge k | X_0 = i) = m \cdot \mathbb{E}(\sigma(i) | X_0 = i) \end{aligned}$$

Since F is finite, $\mathcal{X}^F$ is positive recurrent by theorem 18. Hence we know that $\mathbb{E}(\sigma(i) | X_0 = i) < \infty$, and thus $\mathbb{E}(\tau_i | X_0 = i) < \infty$ which shows that $\mathcal{X}$ is positive recurrent.

□

An often difficult problem is to determine whether a given Markov chain is positive recurrent or not. Concerning this, we now introduce one of the most important criteria for the existence of stationary distributions of Markov chains occuring in queueing theory. It is known as **Foster's criterion**.

Theorem 20 *Let $\mathcal{X}$ denote an irreducible Markov chain with countable state space E and transition matrix P. Further let F denote a finite subset of E. If there is a function $h : E \to \mathbb{R}$ with $\inf\{h(i) : i \in E\} > -\infty$, such that the conditions*

$$\sum_{k\in E} p_{ik} h(k) < \infty \qquad \textit{and} \qquad \sum_{k\in E} p_{jk} h(k) \leq h(j) - \varepsilon$$

hold for some $\varepsilon > 0$ and all $i \in F$ and $j \in E \setminus F$, then $\mathcal{X}$ is positive recurrent.

Proof: Without loss of generality we can assume $h(i) \geq 0$ for all $i \in E$, since otherwise we only need to increase h by a suitable constant. Define the stopping time $\tau_F := \min\{n \in \mathbb{N}_0 : X_n \in F\}$. First we observe that

$$\begin{aligned}
\mathbb{E}(h(X_{n+1}) \cdot 1_{\tau_F > n+1} | X_0, \dots, X_n) &\leq \mathbb{E}(h(X_{n+1}) \cdot 1_{\tau_F > n} | X_0, \dots, X_n) \\
&= 1_{\tau_F > n} \cdot \sum_{k\in E} p_{X_n,k} h(k) \\
&\leq 1_{\tau_F > n} \cdot (h(X_n) - \varepsilon) \\
&= h(X_n) \cdot 1_{\tau_F > n} - \varepsilon \cdot 1_{\tau_F > n}
\end{aligned}$$

holds for all $n \in \mathbb{N}_0$, where the first equality is due to (13). We now proceed with

$$\begin{aligned}
0 &\leq \mathbb{E}(h(X_{n+1}) \cdot 1_{\tau_F > n+1} | X_0 = i) \\
&= \mathbb{E}(\mathbb{E}(h(X_{n+1}) \cdot 1_{\tau_F > n+1} | X_0, \dots, X_n) | X_0 = i) \\
&\leq \mathbb{E}(h(X_n) \cdot 1_{\tau_F > n} | X_0 = i) - \varepsilon \mathbb{P}(\tau_F > n | X_0 = i) \\
&\leq \dots \\
&\leq \mathbb{E}(h(X_0) \cdot 1_{\tau_F > 0} | X_0 = i) - \varepsilon \sum_{k=0}^{n} \cdot \mathbb{P}(\tau_F > k | X_0 = i)
\end{aligned}$$

which holds for all $i \in E \setminus F$ and $n \in \mathbb{N}_0$. For $n \to \infty$ this implies

$$\mathbb{E}(\tau_F | X_0 = i) = \sum_{k=0}^{\infty} \mathbb{P}(\tau_F > k | X_0 = i) \leq h(i)/\varepsilon < \infty$$

for $i \in E \setminus F$. Now the mean return time to the state set F is bounded by

$$\mathbb{E}(\tau_F|X_0 = i) = \sum_{j\in F} p_{ij} + \sum_{j\in E\setminus F} p_{ij}\mathbb{E}(\tau_F + 1|X_0 = j)$$
$$\leq 1 + \varepsilon^{-1} \sum_{j\in E} p_{ij} h(j) < \infty$$

for all $i \in F$, which completes the proof.
□

Notes

Markov chains originate from a series of papers written by A. Markov at the beginning of the 20th century. His first application is given here as exercise 3. However, methods and terminology at that time were very different from today's presentations.

The literature on Markov chains is perhaps the most extensive in the field of stochastic processes. This is not surprising, as Markov chains form a simple and useful starting point for the introduction of other processes.

Textbook presentations are given in Feller [3], Breiman [1], Karlin and Taylor [5], or Çinlar [2], to name but a few. The treatment in Ross [7] contains the useful concept of time–reversible Markov chains. An exhaustive introduction to Markov chains on general state spaces and conditions for their positive recurrence is given in Meyn and Tweedie [6].

Exercise 1 Let $(X_n : n \in \mathbb{N}_0)$ be a family of iid random variables with discrete state space. Show that $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ is a homogeneous Markov chain.

Exercise 2 Let $(X_n : n \in \mathbb{N}_0)$ be iid random variables on $\mathbb{N}_0$ with probabilities $a_i := \mathbb{P}(X_n = i)$ for all $n, i \in \mathbb{N}_0$. The event $X_n > \max(X_0, \ldots, X_{n-1})$ for $n \geq 1$ is called a record at time n. Define T_i as the time of the ith record, i.e. $T_0 := 0$ and $T_{i+1} := \min\{n \in \mathbb{N} : X_n > X_{T_i}\}$ for all $i \in \mathbb{N}_0$. Denote the ith record value by $R_i := X_{T_i}$. Show that $(R_i : i \in \mathbb{N}_0)$ and $((R_i, T_i) : i \in \mathbb{N}_0)$ are Markov chains by determining their transition probabilities.

Exercise 3 Diffusion model by Bernoulli and Laplace
The following is a stochastic model for the flow of two incompressible fluids between two containers: Two boxes contain m balls each. Of these $2m$ balls, b

are black and the others are white. The system is said to be in state i if the first box contains i black balls. A state transition is performed by choosing one ball out of each box at random (meaning here that each ball is chosen with equal probability) and then interchanging the two. Derive a Markov chain model for the system and determine the transition probabilities.

Exercise 4 Let $\mathcal{X}$ denote a Markov chain with $m < \infty$ states. Show that if state j is accessible from state i, then it is accessible in at most $m-1$ transitions.

Exercise 5 Let $p = (p_n : n \in \mathbb{N}_0)$ be a discrete probability distribution and define

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & \cdots \\ & p_0 & p_1 & \ddots \\ & & p_0 & \ddots \\ & & & \ddots \end{pmatrix}$$

with all non–specified entries being zero. Let $\mathcal{X}$ denote a Markov chain with state space $\mathbb{N}_0$ and transition matrix P. Derive an expression (in terms of discrete convolutions) for the transition probabilities $\mathbb{P}(X_{n+m} = j | X_n = i)$ with $n, m \in \mathbb{N}_0$ and $i, j \in \mathbb{N}_0$. Apply the result to the special case of a Bernoulli process (see example 3).

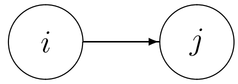
Exercise 6 Prove equation (6).

Exercise 7 Prove the equation $P^n(i,j) = \sum_{k=1}^{n} F_k(i,j) P^{n-k}(j,j)$ for all $n \in \mathbb{N}$ and $i, j \in E$.

Exercise 8 Let $\mathcal{X}$ denote a Markov chain with state space $E = \{1, \ldots, 10\}$ and transition matrix

$$P = \begin{pmatrix}
1/2 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 & 0 & 2/3 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1/3 & 1/3 & 0 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 3/4 & 0 \\
0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 & 1/4 & 0 & 1/4 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 1/3
\end{pmatrix}$$

Reorder the states according to their communication classes and determine the resulting form of the transition matrix as in representation (4). Determine further a transition graph, in which



means that $f_{ij} > 0$.

Exercise 9 Prove equation (7).
Hint: Derive a representation of N_j in terms of the random variables

$$A_n := \begin{cases} 1, & X_n = j \\ 0, & X_n \neq j \end{cases}$$

Exercise 10 Prove corollary 2.

Exercise 11 Prove remark 1.

Exercise 12 A server's up time is k time units with probability $p_k = 2^{-k}$, $k \in \mathbb{N}$. After failure the server is immediately replaced by an identical new one. The up time of the new server is of course independent of the behaviour of all preceding servers.
Let X_n denote the remaining up time of the server at time $n \in \mathbb{N}_0$. Determine the transition probabilities for the Markov chain $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ and determine the stationary distribution of $\mathcal{X}$.

Exercise 13 Let P denote the transition matrix of an irreducible Markov chain $\mathcal{X}$ with discrete state space $E = F \cup F^c$, where $F^c = E \setminus F$. Write P in block notation as

$$P = \begin{pmatrix} P_{FF} & P_{FF^c} \\ P_{F^cF} & P_{F^cF^c} \end{pmatrix}$$

Show that the Markov chain $\mathcal{X}^F$ restricted to the state space F has transition matrix

$$P^F = P_{FF} + P_{FF^c}(I - P_{F^cF^c})^{-1} P_{F^cF}$$

with I denoting the identity matrix on F^c.

Exercise 14 Let $\mathcal{X}$ denote a Markov chain with state space $E = \{0, \ldots, m\}$ and transition matrix

$$P = \begin{pmatrix} p_{00} & p_{01} & & & \\ p_{10} & p_{11} & p_{12} & & \\ & p_{21} & p_{22} & p_{23} & \\ & & \ddots & \ddots & \ddots \\ & & & p_{m,m-1} & p_{mm} \end{pmatrix}$$

where $p_{ij} > 0$ for $|i-j| = 1$. Show that the stationary distribution π of $\mathcal{X}$ is uniquely determined by

$$\pi_n = \pi_0 \cdot \prod_{i=1}^{n} \frac{p_{i-1,i}}{p_{i,i-1}} \qquad \text{and} \qquad \pi_0 = \left(\sum_{j=0}^{m} \prod_{i=1}^{j} \frac{p_{i-1,i}}{p_{i,i-1}} \right)^{-1}$$

for all $n = 1, \ldots, m$.
Use this result to determine the stationary distribution of the Bernoulli–Laplace diffusion model with $b = m$ (see exercise 3).

Exercise 15 Show that the second condition in theorem 20 can be substituted by the condition

$$\sum_{j \in E} p_{ij} h(j) \le h(i) - 1 \qquad \text{for all } i \in E \setminus F.$$

Exercise 16 Show the following complement to theorem 20: Let P denote the transition matrix of a positive recurrent Markov chain with discrete state space E. Then there is a function $h : E \to \mathbb{R}$ and a finite subset $F \subset E$ such that

$$\sum_{j \in E} p_{ij} h(j) < \infty \qquad \text{for all } i \in F, \text{ and}$$

$$\sum_{j \in E} p_{ij} h(j) \le h(i) - 1 \qquad \text{for all } i \in E \setminus F.$$

Hint: Consider the conditional expectation of the remaining time until returning to a fixed set F of states.

Exercise 17 For the discrete, non–negative random walk with transition matrix

$$P = \begin{pmatrix} p_{00} & p_{01} & & & \\ p_{10} & 0 & p_{12} & & \\ & p_{10} & 0 & p_{12} & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

determine the criterion of positive recurrence according to theorem 20.

6 Extension Theorems

Throughout this book, our basic stochastic tools are either sequences of random variables (such as Markov chains or Markov renewal chains) or even uncountable families of random variables (such as Markov processes, renewal processes, or semi–regenerative processes). It is essential for our models that these random variables are dependent, and in fact we define them in terms of conditional probabilities, i.e. via their dependence structure.

It is then an immediate question whether a probability measure $\mathbb{P}$ exists that satisfies all the postulates in the definition of a stochastic sequence or process. This question is vital as it concerns the very existence of the tools we are using.

6.1 Stochastic chains

Let $(S, \mathcal{B})$ denote a measurable space, μ a probability measure on $(S, \mathcal{B})$, and P_n, $n \in \mathbb{N}$, stochastic **kernel**s on $(S, \mathcal{B})$. The latter means that for every $n \in \mathbb{N}$, $P_n : S \times \mathcal{B} \to [0, 1]$ is a function that satisfies
(K1) For every $x \in S$, $P_n(x, .)$ is a probability measure on $(S, \mathcal{B})$.
(K2) For every $A \in \mathcal{B}$, the function $P_n(., A)$ is $\mathcal{B}$–measurable.

Define S^∞ as the set of all sequences $x = (x_n : n \in \mathbb{N}_0)$ with $x_n \in S$ for all $n \in \mathbb{N}_0$. A subset of S^∞ having the form

$$C_{n_1,\ldots,n_k}(A) = \{x \in S^\infty : (x_{n_1}, \ldots, x_{n_k}) \in A\}$$

with $k \in \mathbb{N}$, $n_1 < \ldots < n_k \in \mathbb{N}_0$, and $A \in \mathcal{B}^k$, is called **cylinder** (with coordinates $n_1, \ldots, n_k$ and base A). The set $\mathcal{C}$ of all cylinders in S^∞ forms an algebra of sets. Define $\mathcal{B}^\infty := \sigma(\mathcal{C})$ as the minimal σ–algebra containing $\mathcal{C}$.

Now we can state the extension theorem for sequences of random variables, which is proven in Gikhman and Skorokhod [4], section II.4.

Theorem 21 *There is a probability measure $\mathbb{P}$ on $(S^\infty, \mathcal{B}^\infty)$ satisfying*

$$\mathbb{P}(C_{0,\ldots,k}(A_0 \times \ldots \times A_k)) = \int_{A_0} d\mu(x_0) \int_{A_1} P_1(x_0, dx_1) \ldots \\ \ldots \int_{A_{k-1}} P_{k-1}(x_{k-2}, dx_{k-1})\, P_k(x_{k-1}, A_k) \quad (10)$$

for all $k \in \mathbb{N}_0$, $A_0, \dots, A_k \in \mathcal{B}$. The measure $\mathbb{P}$ is uniquely determined by the system (10) of equations.

The first part of the theorem above justifies our definitions of Markov chains and Markov renewal chains. The second part states in particular that a Markov chain is uniquely determined by its initial distribution and its transition matrix.

Based on this result, we may define a **stochastic chain** with state space S as a sequence $(X_n : n \in \mathbb{N}_0)$ of S–valued random variables which are distributed according to a probability measure $\mathbb{P}$ on $(S^\infty, \mathcal{B}^\infty)$.

7 Conditional Expectations and Probabilities

Let $(\Omega, \mathcal{A}, P)$ denote a probability space and $(S, \mathcal{B})$ a measurable space. A **random variable** is a measurable mapping $X : \Omega \to S$, which means that $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$. In other words, X is a random variable if and only if $X^{-1}(\mathcal{B}) \subset \mathcal{A}$. In stochastic models, a random variable usually gives information on a certain phenomenon, e.g. the number of users in a queue at some specific time.

Consider any real–valued random variable $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$, $\mathcal{B}$ denoting the Borel σ–algebra on $\mathbb{R}$, which is integrable or non–negative. While the random variable X itself yields the full information, a rather small piece of information on X is given by its **expectation**

$$\mathbb{E}(X) := \int_\Omega X \, dP$$

The conditional expectation is a concept that yields a degree of information which lies between the full information X and its expectation $\mathbb{E}(X)$.

To motivate the definition, we first observe that the distribution $P^X = P \circ X^{-1}$ of X is a measure on the sub–σ–algebra $X^{-1}(\mathcal{B})$ of $\mathcal{A}$, i.e. in order to compute

$$P(X \in B) = P^X(B) = \int_{X^{-1}(B)} dP$$

we need to evaluate the measure P on sets

$$A := X^{-1}(B) \in X^{-1}(\mathcal{B}) \subset \mathcal{A}$$

On the other hand, the expectation $\mathbb{E}(X)$ is an evaluation of P on the set $\Omega = X^{-1}(S)$ only. Thus we can say that the expectation employs P only on the trivial

σ–algebra $\{\emptyset, \Omega\}$, while X itself employs P on the σ–algebra $X^{-1}(\mathcal{B})$ generated by X.

Now we take any sub–σ–algebra $\mathcal{C} \subset \mathcal{A}$. According to the Radon–Nikodym theorem there is a random variable $X_0 : \Omega \to S$ with $X^{-1}(\mathcal{B}) = \mathcal{C}$ and

$$\int_C X_0 dP = \int_C X dP \tag{11}$$

for all $C \in \mathcal{C}$. This we call the **conditional expectation** of X under $\mathcal{C}$ and write

$$\mathbb{E}(X|\mathcal{C}) := X_0$$

A conditional expectation is P–almost certainly uniquely determined by (11). Typical special cases and examples are

Example 9 For $\mathcal{C} = \{\emptyset, \Omega\}$, the conditional expectation equals the expectation, i.e. $\mathbb{E}(X|\mathcal{C}) = \mathbb{E}(X)$. For any σ–algebra $\mathcal{C}$ with $X^{-1}(\mathcal{B}) \subset \mathcal{C}$ we obtain $\mathbb{E}(X|\mathcal{C}) = X$.

Example 10 Let I denote any index set and $(Y_i : i \in I)$ a family of random variables. For the σ–algebra $\mathcal{C} = \sigma(\bigcup_{i \in I} Y_i^{-1}(\mathcal{B}))$ generated by $(Y_i : i \in I)$, we write

$$\mathbb{E}(X|Y_i : i \in I) := \mathbb{E}(X|\mathcal{C})$$

By definition we obtain for a σ–algebra $\mathcal{C} \subset \mathcal{A}$, random variables X and Y, and real numbers α and β

$$\mathbb{E}(\alpha X + \beta Y|\mathcal{C}) = \alpha\mathbb{E}(X|\mathcal{C}) + \beta\mathbb{E}(Y|\mathcal{C})$$

For σ–algebras $\mathcal{C}_1 \subset \mathcal{C}_2 \subset \mathcal{A}$ we obtain

$$\mathbb{E}(\mathbb{E}(X|\mathcal{C}_2)|\mathcal{C}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{C}_1)|\mathcal{C}_2) = \mathbb{E}(X|\mathcal{C}_1) \tag{12}$$

Let $\mathcal{C}_1$ and $\mathcal{C}_2$ denote sub–σ–algebras of $\mathcal{A}$, $\mathcal{C} := \sigma(\mathcal{C}_1 \cup \mathcal{C}_2)$, and X an integrable random variable. If $\sigma(X^{-1}(\mathcal{B}) \cup \mathcal{C}_1)$ and $\mathcal{C}_2$ are independent, then

$$\mathbb{E}(X|\mathcal{C}) = \mathbb{E}(X|\mathcal{C}_1)$$

If X and Y are integrable random variables and $X^{-1}(\mathcal{B}) \subset \mathcal{C}$, then

$$\mathbb{E}(XY|\mathcal{C}) = X \cdot \mathbb{E}(Y|\mathcal{C}) \tag{13}$$

Conditional probabilities are special cases of conditional expectations. Define the **indicator function** of a measurable set $A \in \mathcal{A}$ by

$$1_A(x) := \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$$

Such a function is a random variable, since

$$1_A^{-1}(\mathcal{B}) = \{\emptyset, A, A^c, \Omega\} \subset \mathcal{A}$$

with $A^c := \Omega \setminus A$ denoting the complement of the set A. Let $\mathcal{C}$ denote a sub–σ–algebra of $\mathcal{A}$. The conditional expectation of 1_A is called **conditional probability** of A. We write

$$P(A|\mathcal{C}) := \mathbb{E}(1_A|\mathcal{C})$$

Immediate properties of conditional probabilities are

$$0 \leq P(A|\mathcal{C}) \leq 1, \qquad P(\emptyset|\mathcal{C}) = 0, \qquad P(\Omega|\mathcal{C}) = 1$$

$$A_1 \subset A_2 \Longrightarrow P(A_1|\mathcal{C}) \leq P(A_2|\mathcal{C})$$

all of which hold P–almost certainly. For a sequence $(A_n : n \in \mathbb{N})$ of disjoint measurable sets, i.e. $A_n \in \mathcal{A}$ for all $n \in \mathbb{N}$ and $A_i \cap A_j = \emptyset$ for $i \neq j$, we obtain

$$P\left(\bigcup_{n=1}^{\infty} A_n \middle| \mathcal{C}\right) = \sum_{n=1}^{\infty} P(A_n|\mathcal{C})$$

P–almost certainly. Let $X : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ denote a non–negative or integrable random variable and $Y : (\Omega, \mathcal{A}) \to (\Omega', \mathcal{A}')$ a random variable. Then there is a measurable function $g : (\Omega', \mathcal{A}') \to (\mathbb{R}, \mathcal{B})$ with

$$\mathbb{E}(X|Y) = g \circ Y$$

This is P^Y–almost certainly determined by

$$\int_{A'} g \, dP^Y = \int_{Y^{-1}(A')} X \, dP$$

for all $A' \in \mathcal{A}'$. Then we can define the conditional probability of X given $Y = y$ as $g(y)$. We write

$$\mathbb{E}(X|Y = y) := g(y)$$

for all $y \in \Omega'$.

References

[1] L. Breiman. *Probability.* Philadelphia, PA: SIAM, 1968.

[2] E. Çinlar. *Introduction to stochastic processes.* Englewood Cliffs, N. J.: Prentice-Hall, 1975.

[3] W. Feller. *An introduction to probability theory and its applications. Vol. I.* New York etc.: John Wiley and Sons, 1950.

[4] I. Gihman and A. Skorohod. *The theory of stochastic processes I.* Berlin etc.: Springer, 1974.

[5] S. Karlin and H. M. Taylor. *A first course in stochastic processes. 2nd ed.* New York etc.: Academic Press, 1975.

[6] S. Meyn and R. Tweedie. *Markov chains and stochastic stability.* Berlin: Springer-Verlag, 1993.

[7] S. Ross. *Stochastic Processes*. John Wiley & Sons, New York etc., 1983.