Introduction to Stochastic Processes

Lothar Breuer

Contents

Chapter 1

Some general definitions

see notes under
http://www.kent.ac.uk/IMS/personal/lb209/files/notes1.pdf

Chapter 2

Markov Chains and Queues in Discrete Time

2.1 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

$$\mathbb{P}(X_{n+1} = j | X_0 = i_0, \ldots, X_n = i_n) = \mathbb{P}(X_{n+1} = j | X_n = i_n) \tag{2.1}$$

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 (2.1) is referred to as the **Markov property**.

Example 2.1 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 2.2 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 2.3 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 2.4 *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 2.5 *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

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

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 2.5 immediately yields

$$\mathbb{P}\left(X_0 = j_0, X_1 = j_1, \ldots, X_m = j_m\right) = \pi_{j_0} \cdot p_{j_0,j_1} \ldots p_{j_{m-1},j_m} \tag{2.2}$$

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 2.5 and the extension theorem by Tulcea (see appendix 5.2) 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 2.6 *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

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

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 2.6 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

$$\mathbb{P}(\tau \leq n|\mathcal{X}) = \mathbb{P}(\tau \leq n|X_0, \ldots, X_n) \tag{2.3}$$

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 2.7 *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.
□

2.2 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 2.8 *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.
□

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 2.9 Let $\mathcal{X}$ denote a discrete random walk (see example 2.2) with the specification $\pi_1 = p$ and $\pi_{-1} = 1 - p$ for some parameter $0 < p < 1$. Then $\mathcal{X}$ is irreducible.

Example 2.10 The Bernoulli process (see example 2.3) 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 2.11 *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

$$P = \begin{bmatrix} Q & Q_1 & Q_2 & Q_3 & Q_4 & \dots \\ 0 & P_1 & 0 & 0 & 0 & \dots \\ 0 & 0 & P_2 & 0 & 0 & \dots \\ 0 & 0 & 0 & P_3 & 0 & \dots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \end{bmatrix} \tag{2.4}$$

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 2.12 *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

$$f_{ij} := \mathbb{P}(\tau_j < \infty | X_0 = i) = \sum_{k=1}^{\infty} F_k(i,j) \tag{2.5}$$

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 2.12 leads to

$$f_{ij} = p_{ij} + \sum_{h \neq j} p_{ih} f_{hj} \tag{2.6}$$

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 (2.6) is useful for computing the distribution of N_j:

Theorem 2.13 *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 2.7) 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 2.14 *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

$$r_{ij} = \sum_{n=0}^{\infty} P^n(i,j) \tag{2.7}$$

for all $i, j \in E$. The results in theorem 2.13 and corollary 2.14 yield

Corollary 2.15 *For all $i, j \in E$ the relations*

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

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 2.16 *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 $0 < P^m(i,j) \leq 1$ and $0 < P^n(j,i) \leq 1$. The inequalities

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

now imply $r_{jj} < \infty$ because of representation (2.7). According to corollary 2.15 this means that j is transient, too.
If j is recurrent, then the same inequalities lead to

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

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

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

according to formula (2.6), since $f_{ji} = 0$ (otherwise $i \leftrightarrow j$). Thus i is transient, which proves the second statement.
□

Theorem 2.17 *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 2.15 yields $r_{jj} < \infty$. The second equation in the same corollary now implies $r_{ij} < \infty$, which by the representation (2.7) completes the proof.
□

2.3 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 2.18 *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.*

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

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 2.19 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 2.20 Bernoulli process (see example 2.1)
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 2.21 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 2.20):

Theorem 2.22 *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 2.17, 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 2.15), 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 as theorem 4.12) states that

$$\lim_{n \to \infty} \frac{\mathbb{E}(N_i(n) | X_0 = j)}{n} = \frac{1}{m_i} \tag{2.8}$$

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 2.23 *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 2.24 *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

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

The particular statements in the theorem are obvious from theorem 2.16 and the definition of π. The stationarity of π is shown as follows. First we obtain

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

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

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

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 2.25 *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 2.24. Uniqueness of the stationary distribution can be seen as follows. Let π denote the stationary distribution as constructed in theorem 2.24 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 2.26 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 2.27 *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 2.24 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 2.28 *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 (2.8). 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.
□

2.4 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 2.7) 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 2.29 *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 2.24 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 2.30 *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

$$\nu_j = \sum_{k \in F} \nu_k \sum_{n=0}^{\infty} \mathbb{P}(X_n = j, \tau_F > n | X_0 = k) \tag{2.9}$$

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

Proof: Due to theorem 2.29 it suffices to prove equation (2.9) 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 2.24 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 2.24 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.
□

2.5 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 2.31 *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 2.15 that h is recurrent and by irreducibility that the chain is recurrent.
If the chain were null recurrent, then according to the relation in (2.8)

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

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 2.32 *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

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

Since F is finite, $\mathcal{X}^F$ is positive recurrent by theorem 2.31. 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 2.33 *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

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

holds for all $n \in \mathbb{N}_0$, where the first equality is due to (5.3). We now proceed with

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

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

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

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

2.6 The M/M/1 queue in discrete time

Choose any parameters $0 < p, q < 1$. Let the arrival process be distributed as a Bernoulli process with parameter p and the service times $(S_n : n \in \mathbb{N}_0)$ be iid according to the geometric distribution with parameter q.
The geometric service time distribution and the Bernoulli arrival process have been chosen because this simplifies the formulation of the system process in terms of a Markov model due to the following **memoryless property**:

Theorem 2.34 *Let S be distributed geometrically with parameter q, i.e. let $\mathbb{P}(S = k) = (1-q)^{k-1}q$ for all $k \in \mathbb{N}$. Then $\mathbb{P}(S = k | S > k-1) = q$ holds for the conditional distribution, independently of k. Likewise, if Z_n is the nth inter–arrival time of a Bernoulli process with parameter p, then the relation $\mathbb{P}(Z_n = k | Z_n > k-1) = p$ holds, independently of k and n.*

Proof: First the proof for the geometric distribution: For all $k \in \mathbb{N}$, the argument

$$\begin{aligned}\mathbb{P}(S = k|S > k-1) &= \frac{\mathbb{P}(S = k, S > k-1)}{\mathbb{P}(S > k-1)} = \frac{\mathbb{P}(S = k)}{\mathbb{P}(S > k-1)} \\ &= \frac{(1-q)^{k-1}q}{(1-q)^{k-1}} = q\end{aligned}$$

holds, which shows the first statement. For a Bernoulli process, the nth inter–arrival time $Z_n = T_n - T_{n-1}$ is distributed geometrically with parameter p, due to the strong Markov property. This completes the proof for the second statement. □

Thus the memoryless property states that no matter how long a service time or an inter–arrival time has already passed, the probability of a service completion or an arrival at the next time instant is always the same. Hence the system process $\mathcal{Q} = (Q_n : n \in \mathbb{N}_0)$ of the M/M/1 queue in discrete time with arrival process $\mathcal{T}$ and service times S_n can be formulated easily as a homogeneous Markov chain. It has state space $E = \mathbb{N}_0$ and transition probabilities $p_{01} := p$, $p_{00} := 1 - p$, and

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

for $i \geq 1$. Because of the simple state space, the transition matrix can be displayed in the form of a triagonal matrix

$$P = \begin{pmatrix} 1-p & p & 0 & \dots \\ q(1-p) & pq + (1-p)(1-q) & p(1-q) & \ddots \\ 0 & q(1-p) & pq + (1-p)(1-q) & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$$

Since $p, q > 0$, the chain $\mathcal{Q}$ is irreducible. If $p < q$, then $h(n) := n$ defines a function which satisfies the conditions for Foster's criterion, as

$$\begin{aligned}\sum_{k=0}^{\infty} p_{ik} h(k) &= q(1-p) \cdot (i-1) + (qp + (1-q)(1-p)) \cdot i \\ &\quad + p(1-q) \cdot (i+1) \\ &= i - q(1-p) + p(1-q) = i - q + p \leq i - \varepsilon\end{aligned}$$

for all $i \in \mathbb{N}$, with $\varepsilon = q - p > 0$, and $\sum_{k=0}^{\infty} p_{0k} \cdot h(k) = p < \infty$ show. The ratio p/q is called the **load** of the queue. Thus the system process $\mathcal{Q}$ is positive recurrent if the queue load is less than one.
In order to derive a stationary distribution for $\mathcal{Q}$, we first introduce notations $p' := p(1-q)$ and $q' := q(1-p)$. Then we translate the condition $\pi P = \pi$ into the equations

$$\pi_0 = \pi_0(1-p) + \pi_1 q' \tag{2.10}$$
$$\pi_1 = \pi_0 p + \pi_1(1-p-q') + \pi_2 q' \tag{2.11}$$
$$\pi_n = \pi_{n-1} p' + \pi_n(1-(p'+q')) + \pi_{n+1} q' \tag{2.12}$$

for all $n \geq 2$. For the solution, we guess the geometric form

$$\pi_{n+1} = \pi_n \cdot r$$

for all $n \geq 1$, with $r > 0$. Thus equation (2.12) becomes

$$0 = \pi_n p' - \pi_n r(p'+q') + \pi_n r^2 q' = \pi_n \left(p' - r(p'+q') + r^2 q'\right)$$

for all $n \geq 1$, which leads for non–trivial $\pi \neq 0$ to the roots $r = 1$ and $r = p'/q'$ of the quadratic term.
In the first case $r = 1$, we obtain $\pi_{n+1} = \pi_n$ for all $n \geq 1$. This implies $\sum_{j \in E} \pi_j = \infty$ and thus cannot lead to a stationary distribution. Hence in the case $r = 1$ the geometric approach is not successful.
The second root $r = p'/q'$ allows solutions for the other equations (2.10) and (2.11) too. This can be checked as follows: First, the relation

$$\pi_1 = \pi_0 \frac{p}{q'} = \pi_0 \frac{\rho}{1-p}$$

is a requirement from equation (2.10). Then the second equation (2.11) yields

$$\begin{aligned}\pi_2 &= \frac{1}{q'}\left(\pi_1(p'+q') - \pi_0 p\right) = \frac{1}{q'}\left(\frac{p}{q'}(p'+q') - p\right)\pi_0 \\ &= \pi_0 \frac{p}{q'}\left(\frac{p'+q'}{q'} - 1\right) = \pi_1 \frac{p'}{q'}\end{aligned}$$

in accordance with our geometric approach. Now normalization of π leads to

$$1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 \left(1 + \frac{p}{q'} \sum_{n=1}^{\infty} \left(\frac{p'}{q'}\right)^{n-1}\right)$$

from which we obtain

$$\begin{aligned}\pi_0 &= \left(1 + \frac{p}{q'}\sum_{n=1}^{\infty}\left(\frac{p'}{q'}\right)^{n-1}\right)^{-1} = \left(1 + \frac{p}{q'(1-p'/q')}\right)^{-1} \\ &= \left(1 + \frac{p}{q'-p'}\right)^{-1} = (q'-p')(q'-p'+p)^{-1} = \frac{q-p}{q} \\ &= 1-\rho\end{aligned}$$

with $\rho := p/q$, because of $q' - p' = q - p$. Hence the approach $\pi_{n+1} = \pi_n \cdot r$ with $r = p'/q'$ leads to a solution of $\pi P = \pi$.
Note that $r < 1$ if and only if $p < q$. Further, the mean inter–arrival time is $\mathbb{E}(T_1) = 1/p$ and the mean service time is $\mathbb{E}(S_1) = 1/q$. Thus the geometric approach is successful if the so–called **stability condition**

$$\rho = \frac{p}{q} = \frac{\mathbb{E}(S_1)}{\mathbb{E}(T_1)} < 1$$

holds. This condition simply postulates that the mean service time be shorter than the mean inter–arrival time. In this case, the stationary distribution π of $\mathcal{Q}$ has the form

$$\pi_0 = 1-\rho \qquad \text{and} \qquad \pi_n = (1-\rho)\frac{\rho}{1-p}r^{n-1}$$

for all $n \geq 1$. It thus is a modified geometric distribution with parameter $r = p'/q' < 1$.

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 2.37. 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 [5], Breiman [2], Karlin and Taylor [11], or Çinlar [3], to name but a few. The treatment in Ross [19] 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 [15].

Exercise 2.35 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 2.36 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 2.37 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 2.38 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 2.39 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 2.3).

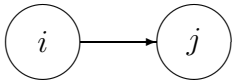
Exercise 2.40 Prove equation (2.6).

Exercise 2.41 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 2.42 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 (2.4). Determine further a transition graph, in which



means that $f_{ij} > 0$.

Exercise 2.43 Prove equation (2.7).
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 2.44 Prove corollary 2.15.

Exercise 2.45 Prove remark 2.26.

Exercise 2.46 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 2.47 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 2.48 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 2.37).

Exercise 2.49 Show that the second condition in theorem 2.33 can be substituted by the condition

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

Exercise 2.50 Show the following complement to theorem 2.33: 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

$$\sum_{j \in E} p_{ij} h(j) < \infty \qquad \text{for all } i \in F, \text{ and}$$
$$\sum_{j \in E} p_{ij} h(j) \leq h(i) - 1 \qquad \text{for all } i \in E \setminus F.$$

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

Exercise 2.51 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 2.33.

Chapter 3

Homogeneous Markov Processes on Discrete State Spaces

In the present chapter we will transfer the discrete time results of the previous chapter to Markov processes in continuous time.

3.1 Definition

Define $T_0 := 0$ and let $(T_n : n \in \mathbb{N})$ denote a sequence of positive real–valued random variables with $T_{n+1} > T_n$ for all $n \in \mathbb{N}_0$ and $T_n \to \infty$ as $n \to \infty$. Further, let E denote a countable state space and $(X_n : n \in \mathbb{N}_0)$ a sequence of E–valued random variables. A process $\mathcal{Y} = (Y_t : t \in \mathbb{R}_0^+)$ in continuous time with

$$Y_t := X_n \qquad \text{for} \qquad T_n \leq t < T_{n+1}$$

is called a **pure jump process**. The variable $H_n := T_{n+1} - T_n$ (resp. X_n) is called the nth **holding time** (resp. the nth **state**) of the process $\mathcal{Y}$. If further $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ is a Markov chain with transition matrix $P = (p_{ij})_{i,j \in E}$ and the variables H_n are independent and distributed exponentially with parameter λ_{X_n} only depending on the state X_n, then $\mathcal{Y}$ is called homogeneous **Markov process** with discrete **state space** E. The chain $\mathcal{X}$ is called the **embedded Markov chain** of $\mathcal{Y}$. As a technical assumption we always agree upon the condition $\hat{\lambda} := \sup\{\lambda_i : i \in E\} < \infty$, i.e. the parameters for the exponential holding times shall be bounded.
An immediate consequence of the definition is that the paths of a Markov process are step functions. The lengths of the holding times are almost certainly strictly

positive, since exponential distributions are zero with probability zero.

Example 3.1 Poisson process
Define $X_n := n$ deterministically. Then $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ is a Markov chain with state space $E = \mathbb{N}_0$ and transition probabilities $p_{n,n+1} = 1$ for all $n \in \mathbb{N}_0$. Let the holding times H_n be distributed exponentially with identical parameter $\lambda > 0$. Then the resulting process $\mathcal{Y}$ as defined in the above definition is a Markov process with state space $\mathbb{N}_0$. It is called **Poisson process** with **intensity** (also: rate or parameter) λ.

Next we want to prove a property similar to the Markov property for Markov chains in discrete time. To this aim, we need to show the **memoryless property** for the exponential distribution, which is the analogue to the memoryless property for geometric distributions in discrete time.

Lemma 3.2 *Let H denote a random variable having an exponential distribution with parameter λ. Then the memoryless property*

$$\mathbb{P}(H > t + s | H > s) = \mathbb{P}(H > t)$$

holds for all time durations $s, t > 0$.

Proof: We immediately check

$$\begin{aligned}\mathbb{P}(H > t + s | H > s) &= \frac{\mathbb{P}(H > t + s, H > s)}{\mathbb{P}(H > s)} = \frac{\mathbb{P}(H > t + s)}{\mathbb{P}(H > s)} \\ &= \frac{e^{-\lambda \cdot (t+s)}}{e^{-\lambda \cdot s}} = e^{-\lambda \cdot t} = \mathbb{P}(H > t)\end{aligned}$$

which holds for all $s, t > 0$.
□

Theorem 3.3 *Let $\mathcal{Y}$ denote a Markov process with discrete state space E. Then the* **Markov property**

$$\mathbb{P}(Y_t = j | Y_u : u \leq s) = \mathbb{P}(Y_t = j | Y_s)$$

holds for all times $s < t$ and states $j \in E$.

Proof: Denote the state at time s by $Y_s = i$. Because of the memoryless property of the exponential holding times, the remaining time in state i is distributed exponentially with parameter λ_i, no matter how long the preceeding holding time has

been. After the holding time in the present state elapses, the process changes to another state j according to the homogeneous Markov chain $\mathcal{X}$. Hence the probability for the next state being j is given by p_{ij}, independently of any state of the process before time s. Now another exponential holding time begins, and thus the past before time s will not have any influence on the future of the process $\mathcal{Y}$.
□
Analogous to the discrete time case, for any two time instances $s < t$ the conditional probabilities $\mathbb{P}(Y_t = j|Y_s = i)$ shall be called the **transition probabilities** from time s to time t. We will now derive a recursion formula for the transition probabilities of a Markov process by conditioning on the number of jumps between time s and time t:

Theorem 3.4 *The transition probabilities of a Markov process $\mathcal{Y}$ are given by*

$$\mathbb{P}(Y_t = j|Y_s = i) = \sum_{n=0}^{\infty} P_{ij}^{(n)}(s,t)$$

for all times $s < t$ and states $i, j \in E$, with

$$P_{ij}^{(0)}(s,t) = \delta_{ij} \cdot e^{-\lambda_i \cdot (t-s)}$$

and recursively

$$P_{ij}^{(n+1)}(s,t) = \int_s^t e^{-\lambda_i \cdot u} \lambda_i \sum_{k \in E} p_{ik} P_{kj}^{(n)}(u,t)\, du$$

for all $n \in \mathbb{N}_0$.

Proof: The above representation follows immediately by conditioning on the number of jumps in $]s,t]$. The expressions $P_{ij}^{(n)}(s,t)$ represent the conditional probabilities that $Y_t = j$ and there are n jumps in $]s,t]$ given that $Y_s = i$. In the recursion formula the integral comprises all times u of a possible first jump along with the Lebesgue density $e^{-\lambda_i \cdot u}\lambda_i$ of this event, after which the probability of n remaining jumps reaching state j at time t is given by $\sum_{k \in E} p_{ik} P_{kj}^{(n)}(u,t)$.
□
For every two time instances $s < t$, define the **transition probability matrix** $P(s,t)$ from time s to time t by its entries

$$P_{ij}(s,t) := \mathbb{P}(Y_t = j|Y_s = i)$$

Using the recursion formula, it is shown by induction on n that the conditional probabilities $P_{ij}^{(n)}(s,t)$ are homogeneous in time, i.e. they satisfy

$$P_{ij}^{(n)}(s,t) = P_{ij}^{(n)}(0,t-s)$$

for all $s < t$. Thus we can from now on restrict the analysis to the transition probability matrices

$$P(t) := P(0,t)$$

with $t \geq 0$. With this notation the Markov property yields the **Chapman–Kolmogorov equations**

$$P(s+t) = P(s)P(t)$$

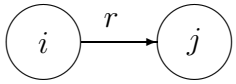
for all time durations $s, t \geq 0$. Thus the family $\{P(t) : t \geq 0\}$ of transition probability matrices forms a semi–group under the composition of matrix multiplication. In particular, we obtain for the neutral element of this semi–group $P(0) = I_E := (\delta_{ij})_{i,j\in E}$ with $\delta_{ij} = 1$ for $i = j$ and zero otherwise.
In order to derive a simpler expression for the transition probability matrices, we need to introduce another concept, which will be called the **generator matrix**. This is defined as the matrix $G = (g_{ij})_{i,j\in E}$ on E with entries

$$g_{ij} := \begin{cases} -\lambda_i \cdot (1 - p_{ii}), & i = j \\ \lambda_i \cdot p_{ij}, & i \neq j \end{cases}$$

for all states $i, j \in E$. In particular, the relation

$$g_{ii} = -\sum_{j \neq i} g_{ij} \tag{3.1}$$

holds for all $i \in E$.
The (i,j)th entry of the generator G is called the **infinitesimal transition rate** from state i to state j. Using these, we can illustrate the dynamics of a Markov process in a directed graph where the nodes represent the states and an edge



means that $g_{ij} = r > 0$. Such a graph is called a **state transition graph** of the Markov process. With the convention $p_{ii} = 0$ the state transition graph uniquely determines the Markov process.

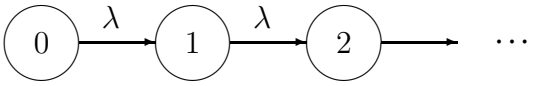


Figure 3.1: Poisson process

Example 3.5 The state transition graph of the Poisson process with intensity λ (see example 3.1) is given by

Theorem 3.6 *The transition probabilities $P_{ij}(t)$ of a Markov process satisfy the systems*

$$\frac{dP_{ij}(t)}{dt} = \sum_{k \in E} P_{ik}(t) g_{kj} = \sum_{k \in E} g_{ik} P_{kj}(t)$$

of differential equations. These are called the **Kolmogorov forward and backward equations**.

Proof: From the representation in theorem 3.4, it follows by induction on the number of jumps that all restricted probabilities $P^{(n)}(t)$ are Lebesgue integrable with respect to t over finite intervals. Since the sum of all $P_{ij}^{(n)}(t)$ is a probability and thus bounded, we conclude by majorized convergence that also $P(t)$ is Lebesgue integrable with respect to t over finite intervals.
Now we can state the recursion

$$P_{ij}(t) = e^{-\lambda_i \cdot t} \cdot \delta_{ij} + \int_0^t e^{-\lambda_i \cdot s} \lambda_i \sum_{k \in E} p_{ik} P_{kj}(t-s)\, ds$$

which results from conditioning on the time s of the first jump from state i. We obtain further

$$P_{ij}(t) = e^{-\lambda_i \cdot t} \cdot \left(\delta_{ij} + \int_0^t e^{+\lambda_i \cdot u} \lambda_i \sum_{k \in E} p_{ik} P_{kj}(u)\, du \right)$$

by substituting $u = t - s$ in the integral. Since $\sum_{k \in E} p_{ik} = 1$ is bounded, we conclude that $P(t)$ is continuous in t. Further, we can differentiate $P(t)$ as given in the recursion and obtain

$$\frac{dP_{ij}(t)}{dt} = -\lambda_i e^{-\lambda_i \cdot t} \cdot \left(\delta_{ij} + \int_0^t f(u)\, du \right) + e^{-\lambda_i \cdot t} \cdot f(t)$$

with f denoting the integrand function. This means nothing else than

$$\begin{aligned}\frac{dP_{ij}(t)}{dt} &= -\lambda_i P_{ij}(t) + \lambda_i \sum_{k\in E} p_{ik} P_{kj}(t) \\ &= -\lambda_i (1 - p_{ii}) \cdot P_{ii}(t) + \sum_{k\neq i} g_{ik} P_{kj}(t)\end{aligned}$$

and thus proves the backward equations. For the forward equations, one only needs to use the Chapman–Kolmogorov equations and apply the backward equations in

$$\begin{aligned}\frac{dP_{ij}(t)}{dt} &= \lim_{h\to 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \lim_{h\to 0} \sum_{k\in E} P_{ik}(t) \frac{P_{kj}(h) - \delta_{kj}}{h} \\ &= \sum_{k\in E} P_{ik}(t) \lim_{h\to 0} \frac{P_{kj}(h) - P_{kj}(0)}{h} = \sum_{k\in E} P_{ik}(t) g_{kj}\end{aligned}$$

which holds for all $i, j \in E$.
□

Theorem 3.7 *The transition probability matrices can be expressed in terms of the generator by*

$$P(t) = e^{G\cdot t} := \sum_{n=0}^{\infty} \frac{t^n}{n!} G^n$$

for all $t \geq 0$, with G^n denoting the nth power of the matrix G.

Proof: First we validate the solution by

$$\frac{d}{dt} e^{G\cdot t} = \frac{d}{dt} \sum_{n=0}^{\infty} \frac{t^n}{n!} G^n = \sum_{n=1}^{\infty} G^n \frac{d}{dt} \frac{t^n}{n!} = \sum_{n=1}^{\infty} G^n \frac{t^{n-1}}{(n-1)!} = G e^{G\cdot t}$$

which holds for all $t \geq 0$. Furthermore, it is obvious that

$$G e^{G\cdot t} = G \sum_{n=0}^{\infty} \frac{t^n}{n!} G^n = \left(\sum_{n=0}^{\infty} \frac{t^n}{n!} G^n \right) G = e^{G\cdot t} G$$

and thus $P(t) = e^{G\cdot t}$ is a solution of Kolmogorov's forward and backward equations.

Now we show uniqueness of the solution. Let $\tilde{P}(t)$ denote another solution of the forward equations. The differential equations with initial condition translate into the integral equations

$$P(t) = I_E + \int_0^t P(u)G\ du \qquad \text{and} \qquad \tilde{P}(t) = I_E + \int_0^t \tilde{P}(u)G\ du$$

Define a norm for matrices $M = (m_{ij})_{i,j\in E}$ on E by

$$\|M\| := \sup\left\{\sum_{j\in E} |m_{ij}| : i \in E\right\}$$

Then $\|G\| \leq 2 \cdot \hat{\lambda}$ and $\|AB\| \leq \|A\| \cdot \|B\|$ for any two matrices A and B on E. Further we obtain

$$\begin{aligned}\left\|P(t) - \tilde{P}(t)\right\| &= \left\|\int_0^t P(u) - \tilde{P}(u)\ du\ G\right\| \\ &\leq \int_0^t \left\|P(u) - \tilde{P}(u)\right\|\ du \cdot \|G\| \qquad (3.2)\\ &\leq \Delta_t \cdot t \cdot \|G\| \qquad (3.3)\end{aligned}$$

with $\Delta_t := \sup\{\|P(u) - \tilde{P}(u)\| : u \leq t\}$, which is finite, since for all $u \geq 0$ we know that $\|P(u)\| = \|\tilde{P}(u)\| = 1$. Plugging the result (3.3) into the right hand of the bound (3.2) again (with time u instead of t), we obtain

$$\left\|P(t) - \tilde{P}(t)\right\| \leq \int_0^t \Delta_t \cdot u \cdot \|G\|\ du \cdot \|G\| = \Delta_t \cdot \frac{t^2}{2} \cdot \|G\|^2$$

Likewise, n–fold repetition of this step achieves the bound

$$\left\|P(t) - \tilde{P}(t)\right\| \leq \Delta_t \cdot \frac{t^n}{n!} \cdot \|G\|^n \leq \Delta_t \cdot \frac{(2\hat{\lambda} \cdot t)^n}{n!}$$

which in the limit $n \to \infty$ yields $0 \leq \left\|P(t) - \tilde{P}(t)\right\| \leq 0$ and consequently $P(t) = \tilde{P}(t)$. As t has been chosen arbitrarily, the statement is proven.
□

Hence the generator of a Markov process uniquely determines all its transition matrices. This can also be seen from the definition, if we agree (without loss of

generality) upon the convention $p_{ii} = 0$ for all $\in E$. Then the parameters for the definition of the Markov process can be recovered by

$$\lambda_i = -g_{ii} \qquad \text{and} \qquad p_{ij} = \frac{g_{ij}}{-g_{ii}}$$

for all $i \neq j \in E$.
However, as in the discrete time case of Markov chains, Markov processes are not completely determined by their transition probability matrices only. The missing link to a complete characterization again is given by the **initial distribution** π with $\pi_i = \mathbb{P}(Y_0 = X_0 = i)$ for all $i \in E$. Then we can express all **finite–dimensional marginal distributions** as in

Theorem 3.8 *For a Markov process $\mathcal{Y}$ with initial distribution π and time instances $0 < t_1 < \ldots < t_n$, $n \in \mathbb{N}$, the equation*

$$\mathbb{P}(Y_{t_1} = j_1, \ldots, Y_{t_n} = j_n) \\ = \sum_{i \in E} \pi_i P_{i,j_1}(t_1) P_{j_1,j_2}(t_2 - t_1) \ldots P_{j_{n-1},j_n}(t_n - t_{n-1})$$

holds for all $j_1, \ldots, j_n \in E$.

The proof is left as an exercise. Thus a Markov process $\mathcal{Y}$ with transition probability matrices $(P(t) : t \geq 0)$ admits a variety of versions depending on the initial distribution π. Any such **version** shall be denoted by $\mathcal{Y}^\pi$.

3.2 Stationary Distribution

From now on we shall convene on the technical assumption

$$\check{\lambda} := \inf\{\lambda_i : i \in E\} > 0$$

which holds for all queueing systems that we will examine. Then a Markov process is called **irreducible**, **transient**, **recurrent** or **positive recurrent** if the defining Markov chain is.
An initial distribution π is called **stationary** if the process $\mathcal{Y}^\pi$ is stationary, i.e. if

$$\mathbb{P}(Y^\pi_{t_1} = j_1, \ldots, Y^\pi_{t_n} = j_n) = \mathbb{P}(Y^\pi_{t_1+s} = j_1, \ldots, Y^\pi_{t_n+s} = j_n)$$

for all $n \in \mathbb{N}$, $0 \leq t_1 < \ldots < t_n$, and states $j_1, \ldots, j_n \in E$, and $s \geq 0$.

Theorem 3.9 *A distribution π on E is stationary if and only if $\pi G = 0$ holds.*

Proof: First we obtain

$$\pi P(t) = \pi e^{G\cdot t} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \pi G^n = \pi I_E + \sum_{n=1}^{\infty} \frac{t^n}{n!} \pi G^n = \pi + \mathbf{0} = \pi$$

for all $t \geq 0$, with $\mathbf{0}$ denoting the zero measure on E. With this, theorem 3.8 yields

$$\begin{aligned}
&\mathbb{P}(Y^\pi_{t_1} = j_1, \ldots, Y^\pi_{t_n} = j_n) \\
&\qquad = \sum_{i\in E} \pi_i P_{i,j_1}(t_1) P_{j_1,j_2}(t_2 - t_1) \ldots P_{j_{n-1},j_n}(t_n - t_{n-1}) \\
&\qquad = \pi_{j_1} P_{j_1,j_2}(t_2 - t_1) \ldots P_{j_{n-1},j_n}(t_n - t_{n-1}) \\
&\qquad = \sum_{i\in E} \pi_i P_{i,j_1}(t_1 + s) P_{j_1,j_2}(t_2 - t_1) \ldots P_{j_{n-1},j_n}(t_n - t_{n-1}) \\
&\qquad = \mathbb{P}(Y^\pi_{t_1+s} = j_1, \ldots, Y^\pi_{t_n+s} = j_n)
\end{aligned}$$

for all times $t_1 < \ldots < t_n$ with $n \in \mathbb{N}$, and states $j_1, \ldots, j_n \in E$. Hence the process $\mathcal{Y}^\pi$ is stationary.
On the other hand, if π is a stationary distribution, then we necessarily obtain $\pi P(t) = \pi e^{G\cdot t} = \pi$ for all $t \geq 0$. As above, this means $\sum_{n=1}^{\infty} \frac{t^n}{n!} \pi G^n = \mathbf{0}$ for all $t \geq 0$, which yields $\pi G = \mathbf{0}$ because of the uniqueness of the zero power series. □
By definition of the generator G and equation (3.1), the equation $\pi G = 0$ is equivalent to an equation system

$$\sum_{i\neq j} \pi_i g_{ij} = -\pi_j g_{jj} \quad \Longleftrightarrow \quad \sum_{i\neq j} \pi_i g_{ij} = \pi_j \sum_{i\neq j} g_{ji} \tag{3.4}$$

for all $j \in E$. This system can be intepreted as follows. We call the value $\pi_i g_{ij}$ **stochastic flow** from state i to state j in equilibrium. Then the above equations mean that the accrued stochastic flow into any state j equals the flow out of this state. Equations (3.4) are called the (global) **balance equations**.

Example 3.10 The generator of the Poisson process with parameter λ (see exam-

ple 3.1) is given by

$$G = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \ddots \\ 0 & 0 & -\lambda & \lambda & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

This process has no stationary distribution, which can be seen as follows. The balance equations for the Poisson process are given by

$$\pi_0\lambda = 0 \qquad \text{and} \qquad \pi_i\lambda = \pi_{i-1}\lambda$$

for all $i \geq 1$. It is immediately evident that these are solvable only by $\pi_i = 0$ for all $i \in E$, which means that there is no stationary distribution π.

The question of existence and uniqueness of a stationary distribution for $\mathcal{Y}$ can be reduced to the same question for $\mathcal{X}$, which we have examined in the preceding chapter:

Theorem 3.11 *Let the underlying Markov chain $\mathcal{X}$ in the definition of the Markov process $\mathcal{Y}$ be irreducible and positive recurrent. Further assume that $\check{\lambda} := \inf\{\lambda_i : i \in E\} > 0$. Then there is a unique stationary distribution for $\mathcal{Y}$.*

Proof: According to theorems 2.25 and 2.18, the transition matrix P of $\mathcal{X}$ admits a unique stationary distribution ν with $\nu P = \nu$. The generator G is defined by $G = \Lambda(P - I_E)$, with $\Lambda = diag(\lambda_i : i \in E)$. Hence the measure $\mu := \nu\Lambda^{-1}$ is stationary for $\mathcal{Y}$. Since $\check{\lambda} > 0$, the measure μ is finite, with total mass bounded by $\check{\lambda}^{-1} < \infty$. Now the normalization

$$\pi_j := \frac{\mu_j}{\sum_{i\in E}\mu_i} = \frac{\nu_j/\lambda_j}{\sum_{i\in E}\nu_i/\lambda_i} \tag{3.5}$$

for all $j \in E$ yields a stationary distribution for $\mathcal{Y}$. This is unique because ν is unique and the construction of π from ν is reversible.
□

Finally we give two important results for the asymptotic behaviour of a Markov process. We call a Markov process **regular** if it satisfies the conditions given in the preceding theorem. If $\mathcal{Y}$ is a regular Markov process, then the limit

$$\lim_{t\to\infty} \mathbb{P}(Y_t = j) = \pi_j \tag{3.6}$$

of the marginal distribution at time t tends to the stationary distribution as t tends to infinity. Further the limit

$$\lim_{t\to\infty} P_{ij}(t) = \pi_j \tag{3.7}$$

holds for all $i, j \in E$ and is independent of i.

Notes

An early text book on Markov processes with discrete state space is Chung [4]. Other classical text book presentation are Karlin and Taylor [11], Breiman [2], or Çinlar [3]. An exposition on non–homogeneous Markov processes on discrete state spaces can be found under the name Markov jump processes in Gikhman and Skorokhod [9, 8].

Exercise 3.12 Consider a population of male and female species. There is an infinitesimal rate $\lambda > 0$ that any male and female produce a single offspring, which will be female with probability p. Determine a Markov process which models the numbers F_t and M_t of female and male species at any time t.

Exercise 3.13 Let X and Y denote two independent random variables which are distributed exponentially with parameters λ and μ, respectively. Prove the following properties:
(a) $X \neq Y$ almost certainly.
(b) The random variable $Z := \min\{X, Y\}$ is distributed exponentially with parameter $\lambda + \mu$.
(c) $\mathbb{P}(X < Y) = \lambda/(\lambda + \mu)$

Exercise 3.14 Let $\mathcal{Y}^{(1)}$ and $\mathcal{Y}^{(2)}$ denote independent Poisson processes with intensities λ_1 and λ_2, respectively. Show that the process $\mathcal{Y} = (Y_t : t \in \mathbb{R}_0^+)$ defined by $Y_t = Y_t^{(1)} + Y_t^{(2)}$ for all $t \geq 0$ is a Poisson process with intensity $\lambda = \lambda_1 + \lambda_2$. The process $\mathcal{Y}$ is called the **superposition** of $\mathcal{Y}^{(1)}$ and $\mathcal{Y}^{(2)}$.

Exercise 3.15 Prove theorem 3.8.

Exercise 3.16 Determine the finite–dimensional marginal distributions for a Poisson process with parameter λ.

Exercise 3.17 Let $\mathcal{Y}$ denote a Poisson process with parameter λ. Given that there is exactly one arrival in the interval $[0, t]$, show that the exact time of the arrival within $[0, t]$ is uniformly distributed.

Exercise 3.18 Verify the Chapman–Kolmogorov equations for a Poisson process.

3.3 First hitting times for Markov processes with a finite state space

As a motivation, we begin with a practical example. Consider the **M/M/c/c+K queue**, which is defined as follows. Arrivals are modelled by a Poisson process with rate $\lambda > 0$. Service times are exponentially distributed with rate $\mu > 0$. There are c servers, and the capacity of the waiting room is K. That means that in total there is room for $c + K$ users in the system including the servers. If upon an arrival the system is filled, i.e. with $c + K$ users already in it, this arriving user is not admitted into the system. In this case we say that the arriving user is lost. Queueing systems with the possibility of such an event are thus called **loss system**s.

The queue described above is a simple Markov process with generator

$$Q = \begin{pmatrix} -\lambda & \lambda & & & & & & & \\ \mu & -\lambda-\mu & \lambda & & & & & & \\ & 2\mu & -\lambda-2\mu & \lambda & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & c\mu & -\lambda-c\mu & \lambda & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & c\mu & -\lambda-c\mu & \lambda & \\ & & & & & & c\mu & -c\mu & \end{pmatrix}$$

up to the first loss (all non–specified entries equal zero).

From a system administrator's point of view, the loss of a user is regarded as a bad event, and thus the question arises naturally how the distribution of the time up to the first loss might be expressed. However, the above description of the queueing process simply ignores loss events, as can be seen from the missing λ entries in the last line of the generator.

In order to include a possible loss event into our model of the queue, we add a

new element to the state space and enlarge the generator as follows:

$$Q' = \begin{pmatrix} -\lambda & \lambda & & & & & & \\ \mu & -\lambda-\mu & \lambda & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & c\mu & -\lambda-c\mu & \lambda & & & \\ & & & \ddots & \ddots & \ddots & & \\ & & & & c\mu & -\lambda-c\mu & \lambda & \\ & & & & & c\mu & -\lambda-c\mu & \lambda \\ 0 & & \dots & & \dots & & & 0 \end{pmatrix}$$

again with all non–specified entries being zero. The first $m = c + K + 1$ states describe the number of users in the system, just as in the former generator Q. But now there is the possibility to enter another state $m + 1$ with rate λ from state m, obviously meaning exactly that a loss has occured. Since we want to observe the system only until the first loss, we choose the loss state $m + 1$ as an absorbing one. Thus all entries in the last line are zero.
Now the system administrator's question can be formulated mathematically as the distribution of the time until the Markov process with generator Q' enters the absorbing state $m + 1$. Exactly this problem is addressed (in a general form) by the concept of a phase–type distribution.

3.3.1 Definition and Examples

Definition 3.19 Let $\mathcal{X} = (X_t : t \geq 0)$ denote an homogeneous Markov process with finite state space $\{1, \dots, m + 1\}$ and generator

$$Q = \begin{pmatrix} T & \eta \\ \mathbf{0} & 0 \end{pmatrix}$$

where T is a square matrix of dimension m, η a column vector and $\mathbf{0}$ the zero row vector of the same dimension. The **initial distribution** of $\mathcal{X}$ shall be the row vector $\tilde{\alpha} = (\alpha, \alpha_{m+1})$, with α being a row vector of dimension m. The first states $\{1, \dots, m\}$ shall be transient, while the state $m + 1$ is absorbing. Let $Z := \inf\{t \geq 0 : X_t = m + 1\}$ be the random variable of the **time until absorption** in state $m + 1$.
The distribution of Z is called **phase–type distribution** (or shortly **PH distribution**) with parameters (α, T). We write $Z \sim PH(\alpha, T)$. The dimension m of T

is called the **order** of the distribution $PH(\alpha, T)$. The states $\{1, \ldots, m\}$ are also called **phase**s, which gives rise to the name phase–type distribution.

Let $\mathbf{1}$ denote the column vector of dimension m with all entries equal to one. The first observations to be derived from the above definition are

$$\eta = -T\mathbf{1} \quad \text{and} \quad \alpha_{m+1} = 1 - \alpha\mathbf{1}$$

These follow immediately from the properties that the row sums of a generator are zero and the sum of a probability vector is one. The vector η is called the **exit vector** of the PH distribution. Now the distribution function and the density of a PH distribution are derived in

Theorem 3.20 *Let $Z \sim PH(\alpha, T)$. Then the distribution function of Z is given by*

$$F(t) := \mathbb{P}(Z \leq t) = 1 - \alpha e^{T \cdot t}\mathbf{1} \tag{3.8}$$

for all $t \geq 0$, and the density function is

$$f(t) = \alpha e^{T \cdot t}\eta \tag{3.9}$$

for all $t > 0$. Here, the function $e^{T \cdot t} := \exp(T \cdot t) := \sum_{n=0}^{\infty} \frac{t^n}{n!} T^n$ denotes a **matrix exponential function**.

Proof: For the Markov process $\mathcal{X}$ with generator Q as given in definition 3.19 the equation

$$P(t) = \exp(Q \cdot t) = \begin{pmatrix} e^{T \cdot t} & \mathbf{1} - e^{T \cdot t}\mathbf{1} \\ \mathbf{0} & 1 \end{pmatrix}$$

holds for the transition matrix $P(t)$ at every time $t \geq 0$. This implies

$$\begin{aligned} F(t) &= \tilde{\alpha} e^{Q \cdot t} e_{m+1} = \alpha_{m+1} + \alpha \cdot (\mathbf{1} - e^{T \cdot t}\mathbf{1}) = \alpha_{m+1} + \alpha\mathbf{1} - \alpha e^{T \cdot t}\mathbf{1} \\ &= 1 - \alpha e^{T \cdot t}\mathbf{1} \end{aligned}$$

with e_{m+1} denoting the $m+1$st canonical base vector. For the density function we obtain

$$f(t) = F'(t) = -\alpha \frac{d}{dt} e^{T \cdot t}\mathbf{1} = -\alpha T e^{T \cdot t}\mathbf{1} = \alpha e^{T \cdot t}(-T\mathbf{1}) = \alpha e^{T \cdot t}\eta$$

which was to be proven.
□
In order to show the versatility of the phase–type concept, we shall give a few examples below. Important characteristics for distributions are their moments. Given a distribution function F, its nth **moment** (if existing) is given by

$$M_n(F) := \int_0^\infty t^n dF(t)$$

Clearly, the first moment is the mean of the distribution. The nth moment of the exponential distribution with parameter λ is given by $M_n = n!/(\lambda^n)$ (see exercises). Another important characteristic is the so–called squared coefficient of variation, defined by

$$C_V(F) := \mathbb{V}ar(F)/\mathbb{E}(F)^2$$

with $\mathbb{V}ar(F)$ denoting the variance of F. For any exponential distribution this equals one. The values of the squared coefficient of variation will explain the names for the hypo- and hyper–exponential distributions introduced below.

Example 3.21 Erlang distribution
A well–known distribution within the family of Gamma distributions is the so–called **Erlang distribution**. An Erlang distribution E_n^λ with n degrees of freedom (or stages) and parameter λ is the distribution of the sum of n exponential random variables with parameter λ. It has the density function

$$f(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t}$$

for all $t \geq 0$. Its interpretation as a succession of n exponential distributions with rate λ each can be illustrated graphically as in

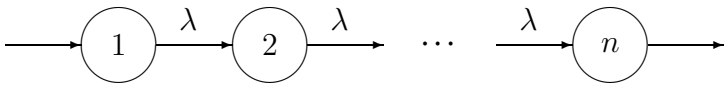


Figure 3.2: Erlang distribution

Here we see that an Erlang distribution can be represented as the holding time in the transient state set $\{1, \ldots, n\}$ of a Markov chain with absorbing state $n + 1$

where the only possible transitions occur from a state k to the next state $k+1$ (for $k = 1, \ldots, n$), with rate λ each. In terms of our definition 3.19, we have a PH representation

$$\alpha = (1, 0, \ldots, 0), \quad T = \begin{pmatrix} -\lambda & \lambda & & \\ & \ddots & \ddots & \\ & & -\lambda & \lambda \\ & & & -\lambda \end{pmatrix} \quad \text{and} \quad \eta = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \lambda \end{pmatrix}$$

with all non–specified entries in T being zero.
The mean of an Erlang distribution with n degrees of freedom and parameter λ is n/λ, while its squared coefficient of variation is $1/n$, i.e. less than one if $n > 1$ (see exercises). This explains the name **hypo–exponential distribution** appearing in the next example.

Example 3.22 Generalized Erlang distribution
A slight generalization of the Erlang distribution is obtained if one admits the exponential stages to have different parameters. Then we talk about a generalized Erlang (or a hypo–exponential) distribution. The representation as a PH distribution results in the figure

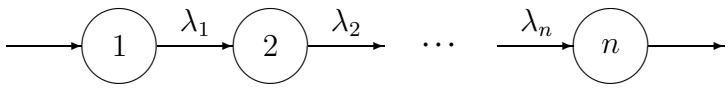


Figure 3.3: Generalized Erlang distribution

and leads to a PH representation

$$\alpha = (1, 0, \ldots, 0), \quad T = \begin{pmatrix} -\lambda_1 & \lambda_1 & & \\ & \ddots & \ddots & \\ & & -\lambda_{n-1} & \lambda_{n-1} \\ & & & -\lambda_n \end{pmatrix} \quad \text{and} \quad \eta = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \lambda_n \end{pmatrix}$$

with all non–specified entries in T being zero. For this family of distributions, a closed formula for the density function is already rather complex.

Example 3.23 Hyper–exponential distribution
A **hyper–exponential distribution** is a finite mixture of $n \in \mathbb{N}$ exponential distributions with different parameters λ_k ($k = 1, \ldots, n$). Its density function is given as

$$f(t) = \sum_{k=1}^{n} q_k \lambda_k e^{-\lambda_k t}$$

with proportions $q_k > 0$ satisfying $\sum_{k=1}^{n} q_k = 1$. A graphical representation of this distribution is

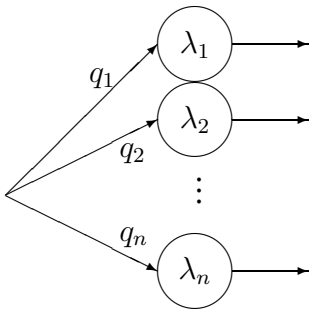


Figure 3.4: Hyper–exponential distribution

This leads to a PH representation by

$$\alpha = (\pi_1, \ldots, \pi_n), \quad T = \begin{pmatrix} -\lambda_1 & & \\ & \ddots & \\ & & -\lambda_n \end{pmatrix} \quad \text{and} \quad \eta = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

with all non–specified entries in T being zero.
The mean of a hyper–exponential distribution is $\sum_{i=1}^{n} \pi_i / \lambda_i$, while its squared coefficient of variation is always larger than one if $n > 1$. This explains the name hyper–exponential distribution.

Example 3.24 Cox distribution
A more complex example of the classical families of distributions are the **Cox distribution**s. These are generalized Erlang distributions with preemptive exit options. A Coxian random variable measures the holding time within the box depicted as

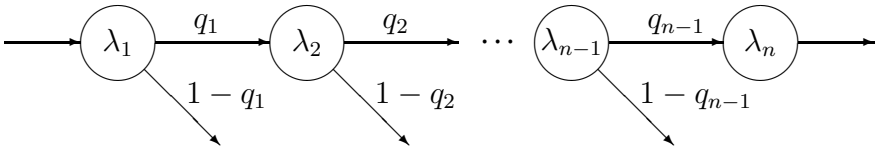


Figure 3.5: Cox distribution

A Cox distribution can be described as a special PH distribution with parameters $\alpha = (1, 0, \ldots, 0)$ and

$$T = \begin{pmatrix} -\lambda_1 & q_1\lambda_1 & & \\ & \ddots & \ddots & \\ & & -\lambda_{n-1} & q_{n-1}\lambda_{n-1} \\ & & & -\lambda_n \end{pmatrix}, \qquad \eta = \begin{pmatrix} (1-q_1)\lambda_1 \\ \vdots \\ (1-q_{n-1})\lambda_{n-1} \\ \lambda_n \end{pmatrix}$$

for which all non–specified entries in T are zero.

As we have seen in example 3.23, the set of transient states in a PH distribution may fall apart into several communication classes. The definition of a phase–type distribution leaves open the possibility of a transient communication class which cannot be entered because the respective initial probabilities contained in the row vector α are zero. Such states are called **superfluous**, since the Markov process $\mathcal{X}$ defining the PH distribution will never be in such a state. As we will obtain the same distribution of the time until absorption if we leave out superfluous states, we shall from now on (unless stated otherwise) assume that there are no superfluous states in the definition of a phase–type distribution.

3.3.2 Closure Properties

A useful advantage of phase–type distributions is the fact that certain compositions of PH distributions result in PH distributions again. This means that the class of PH distributions is closed under these compositions. For PH distributed random variables Z_1 and Z_2 we will show **closure properties** for the compositions $Z_1 + Z_2$ (convolution), $pZ_1 + (1-p)Z_2$ with $p \in [0, 1]$ (mixture), and $\min(Z_1, Z_2)$.

Theorem 3.25 *Let $Z_i \sim PH(\alpha^{(i)}, T^{(i)})$ of order m_i for $i = 1, 2$. Then $Z =$*

$Z_1 + Z_2 \sim PH(\alpha, T)$ *of order* $m = m_1 + m_2$ *with representation*

$$\alpha_k = \begin{cases} \alpha_k^{(1)}, & 1 \leq k \leq m \\ \alpha_{m_1+1}^{(1)} \cdot \alpha_{k-m_1}^{(2)}, & m_1 + 1 \leq k \leq m \end{cases}$$

and

$$T = \begin{pmatrix} T^{(1)} & \eta^{(1)}\alpha^{(2)} \\ \mathbf{0} & T^{(2)} \end{pmatrix}$$

where $\eta^{(1)} = -T^{(1)}1_{m_1}$ *and* $\mathbf{0}$ *denotes a zero matrix of appropriate dimension.*

Proof: By definition, Z_i is the random variable of the time until absorption in a Markov process $\mathcal{X}_i$ with transient states $\{1, \ldots, m_i\}$ and an absorbing state which shall be denoted by e_i in this proof. The transition rates of $\mathcal{X}_i$ within the set of transient states are given by the matrix $T^{(i)}$ and the absorption rates from the transient states to the absorbing state are given by the vector $\eta^{(i)}$.
Then the random variable $Z = Z_1 + Z_2$ is the total time duration of first entering e_1 and then e_2 in the Markov process which is structured as follows:

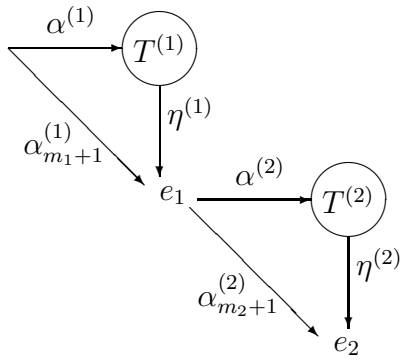


Figure 3.6: Convolution of two PH distributions

Here the point e_1 is not a state of the Markov process described above but only an auxiliary construction aid for a better illustration. In particular there is no holding time in e_1. The only absorbing state in the Markov process constructed above is e_2.
With probability $\alpha_{m_1+1}^{(1)}$ we enter the first absorbing state e_1 immediately, while the vector $\alpha^{(1)}$ contains the probabilities that we first enter the set of transient

states of X_1. In the latter case, the matrix $T^{(1)}$ and then the vector $\eta^{(1)}$ determine the time until the first absorption in e_1.
After having reached e_1, the chain immediately (i.e. with no holding time in e_1) proceeds to the second stage, which is completely analogous to the first. With probability $\alpha^{(2)}_{m_2+1}$ we enter the second absorbing state e_2 immediately, while the vector $\alpha^{(2)}$ contains the probabilities that we first enter the set of transient states of X_2. In the latter case, the matrix $T^{(2)}$ and then the vector $\eta^{(2)}$ determine the time until the first absorption in e_2.
Thus we get to the second absorbing state e_2 immediately with probability $\alpha^{(1)}_{m_1+1} \cdot \alpha^{(2)}_{m_2+1}$. There are transient states $\{1, \ldots, m_1, m_1+1, \ldots, m_1+m_2\}$. The first m_1 of these are reached with probabilities $\alpha_1, \ldots, \alpha_{m_1}$, while the last m_2 of these states can only be reached via an immediate first absorption in e_1 and thus with probabilities $\alpha^{(1)}_{m_1+1} \cdot \alpha^{(2)}_i$ for $i = 1, \ldots, m_2$. This explains the expression for α.
In order to explain the structure of T, we observe first that there is no path from the second set of transient states to the first, whence the lower left entry of T is zero. The diagonal entries of T describe the transition rates within the two sets of transient states, respectively, and thus are given by $T^{(1)}$ and $T^{(2)}$. The only way to get from the first to the second set of transient states is the path via e_1 for which we first need the rates given in $\eta^{(1)}$ and then the probabilities contained in $\alpha^{(2)}$. Hence the upper right entry of T.
□

Theorem 3.26 *Let $Z_i \sim PH(\alpha^{(i)}, T^{(i)})$ of order m_i for $i = 1, 2$, as well as $p \in [0, 1]$. Then $Z = pZ_1 + (1-p)Z_2 \sim PH(\alpha, T)$ of order $m = m_1 + m_2$ with representation*

$$\alpha = (p \cdot \alpha^{(1)}, (1-p) \cdot \alpha^{(2)}) \quad \textit{and} \quad T = \begin{pmatrix} T^{(1)} & \mathbf{0} \\ \mathbf{0} & T^{(2)} \end{pmatrix}$$

where $\mathbf{0}$ denote zero matrices of appropriate dimensions.

Proof: Going along the line of reasoning of the last proof, we observe that Z is equal to Z_1 with probability p and equal to Z_2 with probability $1-p$. Hence we obtain the following construction of a Markov process:

Here, we enter the first set of transient states with probabilities $p \cdot \alpha^{(1)}_i$ for $i = 1, \ldots, m_1$ and the second set with probabilities $(1-p) \cdot \alpha^{(2)}_i$ for phases $i = m_1 + 1, \ldots, m_2$. This explains the expression for α.

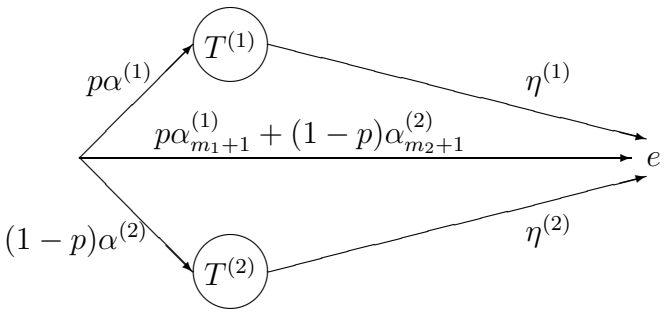


Figure 3.7: Mixture of two PH distributions

From either of these sets we proceed with transition matrices $T^{(i)}$ and exit vectors $\eta^{(i)}$, $i = 1, 2$, in order to reach the absorbing state e. There is no path from one set of transient states to the other, which explains the structure of T.
The absorbing state e can be reached immediately (i.e without entering any transient state) with probability $p\alpha^{(1)}_{m_1+1} + (1-p)\alpha^{(2)}_{m_2+1}$.
□
In order to formulate the next theorem, we first need to define the so–called Kronecker compositions of matrices. Let $A = (a_{ij})$ and $B = (b_{ij})$ denote $n_1 \times m_1$ and $n_2 \times m_2$ matrices, respectively. The **Kronecker product** of A and B is defined as the $(n_1 \cdot n_2) \times (m_1 \cdot m_2)$ matrix $A \otimes B$ with entries

$$(A \otimes B)_{(i_1,i_2),(j_1,j_2)} := a_{i_1,j_1} \cdot b_{i_2,j_2}$$

for all $1 \leq i_k \leq n_k$ and $1 \leq j_k \leq m_k$, $k = 1, 2$. As a block matrix we can write

$$A \otimes B = \begin{pmatrix} a_{11}B & \dots & a_{1m_1}B \\ \vdots & & \vdots \\ a_{n_1 1}B & \dots & a_{n_1 m_1}B \end{pmatrix}$$

If A and B are square matrices, i.e. $n_k = m_k$ for $k = 1, 2$, then the **Kronecker sum** $A \oplus B$ of the matrices A and B is defined as

$$A \oplus B := A \otimes I_2 + I_1 \otimes B$$

with I_k denoting the $n_k \times n_k$ identity matrix for $k = 1, 2$.

Example 3.27 Let $n_1 = m_1 = 1$ and $n_2 = m_2 = 2$. If $A = -\lambda$ and

$$B = \begin{pmatrix} -\mu & \mu \\ 0 & -\mu \end{pmatrix}, \quad \text{then} \quad A \oplus B = \begin{pmatrix} -(\lambda+\mu) & \mu \\ 0 & -(\lambda+\mu) \end{pmatrix}$$

is an explicit expression for the Kronecker sum of A and B.

Theorem 3.28 *Let $Z_i \sim PH(\alpha^{(i)}, T^{(i)})$ of order m_i for $i = 1, 2$ and define $Z = \min(Z_1, Z_2)$. Then $Z \sim PH(\alpha, T)$ of order $m = m_1 \cdot m_2$ with representation*

$$\alpha = \alpha^{(1)} \otimes \alpha^{(2)} \quad \textit{and} \quad T = T^{(1)} \oplus T^{(2)}$$

in terms of the Kronecker compositions.

Proof: For $i = 1, 2$, the random variables Z_i are the times until absorption in the Markov processes $\mathcal{X}_i = (X_t^{(i)} : t \geq 0)$ where the initial distributions for the transient states are $\alpha^{(i)}$ and the transition rates among the transient states are given by $T^{(i)}$. Thus we can determine Z if we start running $\mathcal{X}_1$ and $\mathcal{X}_2$ concurrently and stop whenever the first of the two processes enters the absorbing state. We will show that the two–dimensional Markov process $\mathcal{X}$ depicted as in the figure below has the same time until absorption as the first absorption of the concurrent processes $\mathcal{X}_1$ and $\mathcal{X}_2$.

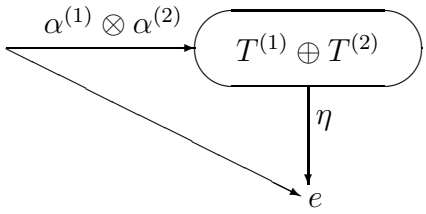


Figure 3.8: Superposition of two PH distributions

The state space of $\mathcal{X}$ shall be

$$E = \{(i, j) : 1 \leq i \leq m_1, 1 \leq j \leq m_2\} \cup \{e\}$$

where e is the absorbing state and all other states are transient. We will keep in mind the interpretation that $X_t = (i, j)$ means $X_t^{(1)} = i$ and $X_t^{(2)} = j$ for all transient states i, j. The exit vector η of dimension $m_1 \cdot m_2$ has entries $\eta_{ij} = \eta_i^{(1)} + \eta_j^{(2)}$ for all $i \leq m_1$ and $j \leq m_2$.

Since we start the processes $\mathcal{X}_1$ and $\mathcal{X}_2$ independently, we clearly have an initial distribution

$$\mathbb{P}(X_0 = (i,j)) = \mathbb{P}(X_0^{(1)} = i) \cdot \mathbb{P}(X_0^{(2)} = j) = \alpha_i^{(1)} \cdot \alpha_j^{(2)}$$

which explains the expression for α. If the process $\mathcal{X}$ is in state (i,j), this means that the exponential holding times in the states i (for $\mathcal{X}_1$) and j (for $\mathcal{X}_2$) are running concurrently. According to lemma 3.13 they almost certainly do not stop at the same time instant. Thus $\mathcal{X}$ has only state transitions from (i,j) to either (i,h) or (k,j). These occur with transition rates $T_{jh}^{(2)}$ or $T_{ik}^{(1)}$, respectively, if h and k are transient states. According to lemma 3.13 the holding time in state (i,j) is exponential with parameter $-(T_{ii}^{(1)} + T_{jj}^{(2)})$. This explains the structure of T. The values of η are readily verified by

$$\begin{aligned}\eta_{ij} &= -(T\mathbf{1}_{m_1 m_2})_{(i,j)} = -\left(T_{ii}^{(1)} + T_{jj}^{(2)} + \sum_{h\neq j} T_{jh}^{(2)} + \sum_{k\neq i} T_{ik}^{(1)}\right) \\ &= -\sum_{k=1}^{m_1} T_{ik}^{(1)} - \sum_{h=1}^{m_2} T_{jh}^{(2)} = \eta_i^{(1)} + \eta_j^{(2)}\end{aligned}$$

Thus we can see that $Z = \min(Z_1, Z_2)$ is the time until absorption in the Markov process $\mathcal{X}$.
□
Example 3.21 and theorem 3.26 already suffice to prove the following powerful

Theorem 3.29 *The class of phase–type distributions is dense (in terms of weak convergence) within the class of all distributions on* $\mathbb{R}_0^+$.

Proof: Let $F : \mathbb{R}_0^+ \to [0,1]$ denote any non–negative distribution function. Since it is bounded, monotone and right–continuous, we can approximate F by a step function G with countably many jumps at $(t_n : n \in \mathbb{N}_0)$, where $t_n < t_{n+1}$ for all $n \in \mathbb{N}$. The error $\varepsilon > 0$ of approximation can be chosen arbitrarily small such that $|F(t) - G(t)| < \varepsilon$ holds for all $t \geq 0$.
If $t_0 = 0$, i.e. if there is a jump of G at zero, we can write

$$G = p_0 \cdot \delta_0 + (1 - p_0) \cdot \tilde{G}$$

with $p_0 = G(0)$ and $\tilde{G} = (G - G(0))/(1 - p_0)$. The Dirac distribution function δ_0 is a phase–type distribution with $m = 0$ and $\alpha_{m+1} = 1$. In view of example 3.21

and theorem 3.26, it now suffices to show that we can approximate the function $\tilde{G}$ by a finite mixture of Erlang distributions. First we find a truncation point T of $\tilde{G}$ such that $G(T) > 1 - \varepsilon$. Then there is a number $N \in \mathbb{N}$ such that $t_n > T$ for all $n > N$. Thus $\tilde{G}$ can be approximated by

$$H = \sum_{n=1}^{N-1} (\tilde{G}(t_n) - \tilde{G}(t_{n-1})) \cdot \delta_{t_n} + (1 - \tilde{G}(t_N)) \cdot \delta_{t_N}$$

with an error bounded by ε.
For every $n = 1, \ldots, N$ we approximate the Dirac distribution δ_{t_n} by a suitable Erlang distribution. This possible because of the following argument: The variance of an Erlang distribution $E_k^{k\lambda}$ of order k with parameter $k \cdot \lambda$ is given by $(k \cdot \lambda^2)^{-1}$ (see exercises) and thus tends to zero as k grows larger. Since the mean of such an Erlang distribution is $1/\lambda$ (see exercises), Chebyshev's inequality tells us that the sequence $(E_k^{k\lambda} : k \in \mathbb{N})$ converges in probability (and hence weakly) towards δ_{t_n} if we chose $\lambda = 1/t_n$. This means that there is a number $K \in \mathbb{N}$ such that the distribution function H_n of an E_K^{K/t_n} distribution satisfies $|H_n(t) - \delta_{t_n}(t)| < \varepsilon$ for all $t \geq 0$.
If we pursue the above approximation method for every $n = 1, \ldots, N$ and define

$$\tilde{H} = \sum_{n=1}^{N-1} (\tilde{G}(t_n) - \tilde{G}(t_{n-1})) \cdot H_n + (1 - \tilde{G}(t_N)) \cdot H_N$$

then we obtain an approximation bound $|H - \tilde{H}| < \varepsilon$. According to example 3.21 and theorem 3.26 the distribution $\tilde{H}$ is phase–type.
In summary, we have approximated F by $p_0 \cdot \delta_0 + (1 - p_0) \cdot \tilde{H}$ with an approximation bound of $3 \cdot \varepsilon$. This proves our statement.
□

Notes

Phase–type distributions have been introduced in Neuts [16] as a generalization of the Erlang and hyper–exponential distribution. A classical introduction to phase–type distributions is given in Neuts [17]. Statistical methods for fitting PH distributions are given in Asmussen et al. [1]. Phase–type distributions with infinitely many phases are introduced in Shi et al. [20], and Shi and Liu [21].
The name of a superfluous state as well as the motivating example at the beginning of the section have been taken from Latouche and Ramaswami [14].

The proofs of the closure properties have been chosen to be as constructive and illustrating as possible. More classical proofs via comparison of the Laplace transforms can be found in Neuts [17].

Exercise 3.30 Show that the density $f(t)$ of a phase–type distribution function is strictly positive for all $t > 0$.

Exercise 3.31 Compute the Laplace–Stieltjes transform, all moments as well as the squared coefficient of variation for the exponential, the Erlang, and the hyper–exponential distribution.

Exercise 3.32 Use the results from exercise 3.31 in order to show that the Erlang distribution with parameter $n\lambda$ and n degrees of freedom is the n–fold convolution of an exponential distribution with parameter $n\lambda$. Employ this result for a simple proof of formula (4.2).

Exercise 3.33 Consider two machines running independently at the same time. The one has a life time which is distributed as a generalized Erlang with two stages and parameters λ_1 and λ_2. The other machine's life time has a hyper–exponential distribution with density $f(t) = p \cdot \mu_1 e^{-\mu_1 t} + (1-p) \cdot \mu_2 e^{-\mu_2 t}$. As soon as a machine fails, it is given to repair which takes an exponentially distributed amount of time with parameter κ. After repair it starts working immediately. Determine the distribution of the time until both machines are broken down.

Exercise 3.34 Consider the M/PH/k queue with Poisson input and phase–type service time distribution for all its k servers. Derive a description for this queue in terms of a $(k+1)$–dimensional Markov process.

Exercise 3.35 Find the stationary distribution for the M/PH/1 queue.

Chapter 4

Renewal Theory

4.1 Renewal Processes

Be $(X_n : n \in \mathbb{N}_0)$ a sequence of independent positive random variables, and assume that $(X_n : n \in \mathbb{N})$ are identically distributed. Define the sequence $\mathcal{S} = (S_n : n \in \mathbb{N})$ by $S_1 := X_0$ and $S_{n+1} := S_n + X_n$ for all $n \in \mathbb{N}$. The random variable S_n, with $n \in \mathbb{N}$, is called the nth **renewal time**, while the time duration X_n is called the nth **renewal interval**. Further define the random variable of the number of renewals until time t by

$$N_t := \max\{n \in \mathbb{N} : S_n \leq t\}$$

for all $t \geq 0$ with the convention $\max \emptyset = 0$. Then the continuous time process $\mathcal{N} = (N_t : t \in \mathbb{R}_0^+)$ is called a **renewal process**. The random variable X_0 is called the **delay** of $\mathcal{N}$. If X_0 and X_1 have the same distribution, then $\mathcal{N}$ is called an **ordinary renewal process**.
We will always assume that $\mathbb{P}(X_1 = 0) = 0$ and $m := \mathbb{E}(X_1) < \infty$ is finite. The strong law of large numbers implies that $S_n/n \to m$ with probability one as $n \to \infty$. Hence $S_n < t$ cannot hold for infinitely many n and thus N_t is finite with probability one. By standard notation we will write

$$G(x) := \mathbb{P}(X_0 \leq x) \qquad \text{and} \qquad F(x) := \mathbb{P}(X_1 \leq x)$$

for all $x \in \mathbb{R}_0^+$.

Example 4.1 A light bulb has been installed at time zero. After a duration X_0, it will go out of order. We assume that it will be immediately replaced by a new light

bulb at time $S_1 = X_0$. Assume that the new light bulb is of a type identical to the old one. Then the duration X_1 until it goes out of order is distributed identically to X_0. Of course, the life times of the light bulbs are independent from one another. Keeping up this rechangement policy over time, the number N_t of used light bulbs until time t forms an ordinary renewal process.

Remark 4.2 A Poisson process with intensity λ (see example 3.1) is an ordinary renewal process with $F(x) = G(x) = 1 - e^{-\lambda x}$, i.e. the renewal intervals have an exponential distribution. Thus a renewal process can be seen as a generalization of the Poisson process with respect to the distribution of the renewal intervals.

In order to derive an expression for the distribution and the expectation of N_t at any time t, we need to introduce the concept of **convolution**s of a non–negative function and a distribution function. Let F denote a distribution function on $\mathbb{R}_0^+$ and $g : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ a Lebesgue–measurable function which is bounded on all finite intervals $[0, t]$ with $t \geq 0$. Then the function defined by

$$F * g(t) := \int_0^t g(t-u)\, dF(u)$$

for all $t \in \mathbb{R}$ is called the convolution of F and g. In particular, the definition of a convolution applies if g is a distribution function. As an exercise the reader can prove

Theorem 4.3 *For any distribution functions F and G as well as non–negative Lebesgue–measurable functions $(g_n : n \in \mathbb{N})$ on $\mathbb{R}_0^+$, the following properties hold:*
(1) The convolution $F * G$ is a distribution function on $\mathbb{R}_0^+$.
(2) $F * G = G * F$
(3) $F * \sum_{n=1}^{\infty} g_n = \sum_{n=1}^{\infty} F * g_n$
(4) The Dirac measure δ_0 on 0 with distribution function I_0, which is defined by $I_0(t) := 1$ for all $t \geq 0$ and $I_0(t) := 0$ otherwise, is neutral in regard to convolutions, i.e. $I_0 * G = G$ for all distribution functions G.
(5) If the random variables X and Y are independent and distributed according to F and G, respectively, then $\mathbb{P}(X + Y \leq t) = F * G(t)$ for all $t \geq 0$.
(6) $F * (G * g) = (F * G) * g$

Let F denote any distribution function for a real–valued random variable. Define the **convolutional power**s by $F^{*1} := F$ and recursively $F^{*n+1} := F^{*n} * F$ for all

$n \in \mathbb{N}$. Because of property (4) in the above theorem, we define $F^{*0} := I_0$ for every distribution function F.
Now denote the distribution function of the random variable X_1 (and hence of all X_n with $n \geq 1$) and X_0 by F and G, respectively. Since the random variables $(X_n : n \in \mathbb{N})$ are iid, part (5) of the above theorem yields for all $n \in \mathbb{N}_0$ the relation $\mathbb{P}(N_t \geq n) = \mathbb{P}(S_n \leq t) = G * F^{*n-1}(t)$ and thus we obtain $\mathbb{P}(N_t = 0) = 1 - G(t)$ and

$$\mathbb{P}(N_t = n) = \mathbb{P}(S_n \leq t) - \mathbb{P}(S_{n+1} \leq t) = G * F^{*n-1}(t) - G * F^{*n}(t)$$

for $n \geq 1$. The expectation of N_t is given by

$$\mathbb{E}(N_t) = \sum_{n=1}^{\infty} \mathbb{P}(N_t \geq n) = \sum_{n=1}^{\infty} \mathbb{P}(S_n \leq t) = G * \sum_{n=0}^{\infty} F^{*n}(t) \tag{4.1}$$

for all $t \geq 0$ (for the first equality see Exercise 4.19). The rate of growth of a renewal process is described by

Theorem 4.4 *Let $\mathcal{N} = (N_t : t \geq 0)$ denote a renewal process with renewal intervals having mean length $m < \infty$. Then*

$$\lim_{t\to\infty} \frac{N_t}{t} = \frac{1}{m}$$

holds with probability one.

Proof: By definition of N_t, the inequalities $S_{N_t} \leq t \leq S_{N_t+1}$ hold with probability one for all times t. Dividing these by N_t and using the strong law of large numbers, we obtain

$$\begin{aligned} m &= \lim_{n\to\infty} \frac{S_n}{n} = \lim_{t\to\infty} \frac{S_{N_t}}{N_t} \\ &\leq \lim_{t\to\infty} \frac{t}{N_t} \\ &\leq \lim_{t\to\infty} \left(\frac{S_{N_t+1}}{N_t+1} \cdot \frac{N_t+1}{N_t} \right) = \lim_{n\to\infty} \frac{S_{n+1}}{n+1} \cdot \lim_{n\to\infty} \frac{n+1}{n} = m \cdot 1 \end{aligned}$$

which proves the statement.
□
Because of this theorem, the inverse $1/m$ of the mean length of a renewal interval is called the **rate** of the renewal process. It describes the asymptotic rate at which renewals occur.

Example 4.5 Regarding a Poisson process $\mathcal{N} = (N_t : t \geq 0)$ with intensity $\lambda > 0$, it can be shown that

$$\mathbb{P}(N_t = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \tag{4.2}$$

for all $t \geq 0$ and $n \in \mathbb{N}_0$. The expectation of N_t is given by $\mathbb{E}(N_t) = \lambda \cdot t$.
Thus a Poisson process with intensity λ has at time t a Poisson distribution with parameter $\lambda \cdot t$. Moreover, the intensity λ is also the rate of the Poisson process, since a mean renewal interval has length $1/\lambda$.
Given an observed stream of events (e.g. job requests at a server) over some time interval of length t, we can count the number $N(t)$ of events that have occurred in this interval. If we want to model such event streams by a Poisson process, then we need to find a statistical estimator for the intensity λ. Now theorem 4.4 states that the fraction $N(t)/t$ comes close to λ for large interval lengths t. Thus a consistent statistical estimator for the intensity λ is given by $\hat{\lambda} = N(t)/t$.

Example 4.6 There is a discrete–time analogue of the Poisson process, which is called **Bernoulli process**. This is an ordinary renewal process with renewal intervals that have a geometric distribution. Given a parameter $p \in]0, 1[$, the length of the renewal intervals is distributed as $\mathbb{P}(X_1 = n) = p \cdot (1-p)^{n-1}$ for $n \in \mathbb{N}$.

4.2 Renewal Function and Renewal Equations

The function defined by $R(t) := \sum_{n=1}^{\infty} F^{*n}(t)$ for all $t \geq 0$ is called the **renewal function** of the process $\mathcal{N}$. The renewal function will play a central role in renewal theory. First we need to show that it remains finite:

Theorem 4.7 *If $F(0) < 1$, then $R(t) = \sum_{n=1}^{\infty} F^{*n}(t) < \infty$ for all $t \geq 0$.*

Proof: Since $F(0) < 1$ and F is continuous to the right, there is a number $\alpha > 0$ such that $F(\alpha) < 1$. Fix any $t \geq 0$ and choose $k \in \mathbb{N}$ such that $k \cdot \alpha > t$. Then $F^{*k}(t) \leq 1 - (1 - F(\alpha))^k =: 1 - \beta$ with $0 < \beta < 1$. Thence we obtain the bound $F^{*mk}(t) \leq (1-\beta)^m$ for any $m \in \mathbb{N}$. Since $F(0-) = 0$, we can use $F^{*n}(t) \geq F^{*h}(t)$ for all $n < h \in \mathbb{N}$. Putting these bounds together, we obtain

$$R(t) = \sum_{n=1}^{\infty} F^{*n}(t) \leq k \cdot \sum_{m=0}^{\infty} F^{*mk}(t) \leq k \cdot \sum_{m=0}^{\infty} (1-\beta)^m = \frac{k}{\beta} < \infty$$

since $\beta > 0$.
□

Theorem 4.8 *An ordinary renewal process is uniquely determined by its renewal function.*

Proof: First we take the Laplace–Stieltjes transform (LST, see appendix 5.3) on both sides of the equation $R(t) = \sum_{n=1}^{\infty} F^{*n}(t)$. This yields

$$\tilde{R}(s) = \sum_{n=1}^{\infty} \widetilde{F^{*n}}(s) = \tilde{F}(s) \cdot \sum_{n=0}^{\infty} (\tilde{F}(s))^n = \frac{\tilde{F}(s)}{1 - \tilde{F}(s)} \tag{4.3}$$

for $s > 0$, or

$$\tilde{F}(s) = \frac{\tilde{R}(s)}{1 + \tilde{R}(s)}$$

and thus determines the LST $\tilde{F}(s)$ of F uniquely in terms of $\tilde{R}(s)$. Now uniqueness of the LST yields the statement.
□
For an ordinary renewal process we can derive an implicit integral equation for the renewal function, which is known as a renewal equation. Note that for an ordinary renewal process $\mathbb{E}(N_t) = R(t)$ for all times t (see (4.1) with $G = F$). Hence the function R is increasing. If we condition upon the length x of the first renewal interval X_0, we obtain

$$\mathbb{E}(N_t) = \int_0^{\infty} \mathbb{E}(N_t | X_0 = x)\, dF(x)$$

Since $\mathbb{E}(N_t|X_0 = x) = 1 + R(t - x)$ for $t \geq x$ and $\mathbb{E}(N_t|X_0 = x) = 0$ for $t < x$, we can simplify this equation to

$$R(t) = \int_0^t (1 + R(t - x))\, dF(x) = F(t) + \int_0^t R(t - x)\, dF(x)$$

for all $t \geq 0$. A **renewal equation** is the generalized form

$$g(t) = h(t) + \int_0^t g(t - x)\, dF(x), \qquad t \geq 0 \tag{4.4}$$

where a function h on $[0, \infty[$ and a distribution function F on $[0, \infty[$ are given and the function g on $[0, \infty[$ is unknown. The solution is given in

Theorem 4.9 *The unique solution g to equation (4.4) is given by*

$$g(t) = \int_0^t h(t-x)\, dR(t) + h(t)$$

where $R(t) = \sum_{n=1}^{\infty} F^{*n}(t)$ denotes the renewal function for F.

Proof: Equation (4.4) can be written as $g = h + g * F$. Because of the definition $R = \sum_{n=1}^{\infty} F^{*n}$ we obtain

$$F * (R * h + h) = F * h + \sum_{n=1}^{\infty} F^{*n+1} * h = \sum_{n=1}^{\infty} F^{*n} * h = R * h$$

which shows that $g = R * h + h$ is indeed a solution of (4.4).
Let g' denote another solution and define the function

$$\delta := g' - R * h - h$$

Then (4.4) implies $\delta = F * \delta$ and thus $\delta = F^{*n} * \delta$ for all $n \in \mathbb{N}$. Since $R(t) < \infty$ for any fixed $t \geq 0$, we infer that $F^{*n} \to 0$ as $n \to \infty$. Hence $\delta(t) = 0$ for all $t \geq 0$, which completes the proof.
□

4.3 Renewal Theorems

In order to present the most powerful results of renewal theory, it will be useful to introduce stopping times and Wald's lemma. Recall from (2.3) that a random variable S with values in $\mathbb{N}_0 \cup \{\infty\}$ is called a stopping time for the sequence $\mathcal{X} = (X_0 : n \in \mathbb{N}_0)$ if

$$\mathbb{P}(S \leq n | \mathcal{X}) = \mathbb{P}(S \leq n | X_0, \ldots, X_n) \tag{4.5}$$

holds for all $n \in \mathbb{N}_0$.

Lemma 4.10 *For a renewal process $\mathcal{N}$ with delay X_0 and renewal intervals $(X_n : n \in \mathbb{N})$, the random variable N_t is a stopping time for the sequence $(X_n : n \in \mathbb{N}_0)$.*

Proof: This follows from the observation that $N_t = k$ is equivalent to

$$\sum_{n=0}^{k-1} X_n \leq t < \sum_{n=0}^{k} X_n$$

which implies that the event $N_t \leq k$ depends only on $X_0, \ldots, X_k$.
□

Lemma 4.11 Wald's Lemma
Be $\mathcal{X} = (X_n : n \in \mathbb{N}_0)$ a sequence of stochastically independent positive random variables with the same expectation $\mathbb{E}(X_n) = m$ for all $n \in \mathbb{N}$. The expectations $\mathbb{E}(X_0)$ and $\mathbb{E}(X_1)$ shall be finite. Further be S a stopping time of the sequence $\mathcal{X}$ with $\mathbb{E}(S) < \infty$. Then

$$\mathbb{E}\left(\sum_{n=0}^{S} X_n\right) = \mathbb{E}(X_0) + \mathbb{E}(S) \cdot m$$

Proof: For all $n \in \mathbb{N}_0$ define the random variables $I_n := 1$ on the set $\{S \geq n\}$ and $I_n := 0$ else. Then $\sum_{n=0}^{S} X_n = \sum_{n=0}^{\infty} I_n X_n$ and hence

$$\mathbb{E}\left(\sum_{n=0}^{S} X_n\right) = \mathbb{E}\left(\sum_{n=0}^{\infty} I_n X_n\right) = \sum_{n=0}^{\infty} \mathbb{E}(I_n X_n)$$

by monotone convergence, as I_n and X_n are non–negative. S being a stopping time for $\mathcal{X}$, we obtain by definition $\mathbb{P}(S \geq 0) = 1$, and further

$$\mathbb{P}(S \geq n|\mathcal{X}) = 1 - \mathbb{P}(S \leq n-1|\mathcal{X}) = 1 - \mathbb{P}(S \leq n-1|X_0, \ldots, X_{n-1})$$

for all $n \in \mathbb{N}$. Since the X_n are independent, I_n and X_n are independent, too, which implies $\mathbb{E}(I_0 X_0) = \mathbb{E}(X_0)$ and

$$\mathbb{E}(I_n X_n) = \mathbb{E}(I_n) \cdot \mathbb{E}(X_n) = \mathbb{P}(S \geq n) \cdot m$$

for all $n \in \mathbb{N}$. Now the relation $\sum_{n=1}^{\infty} \mathbb{P}(S \geq n) = \mathbb{E}(S)$ yields

$$\begin{aligned}\mathbb{E}\left(\sum_{n=0}^{S} X_n\right) &= \sum_{n=0}^{\infty} \mathbb{E}(I_n X_n) = \mathbb{E}(X_0) + \sum_{n=1}^{\infty} \mathbb{P}(S \geq n) \cdot m \\ &= \mathbb{E}(X_0) + \mathbb{E}(S) \cdot m\end{aligned}$$

□

Theorem 4.12 Elementary Renewal Theorem
Be $\mathcal{N}$ a renewal process with renewal intervals $(X_n : n \in \mathbb{N})$ and mean renewal time $\mathbb{E}(X_1) = m > 0$. Assume further that the mean delay is finite, i.e. $\mathbb{E}(X_0) < \infty$. Then for the counting function N_t the limit

$$\lim_{t\to\infty} \frac{\mathbb{E}(N_t)}{t} = \frac{1}{m}$$

holds, with the convention $1/\infty := 0$.

Proof: For every $t \geq 0$, the bound $t < \sum_{n=0}^{N_t} X_n$ holds almost surely. By Wald's lemma, this implies

$$t < \mathbb{E}\left(\sum_{n=0}^{N_t} X_n\right) = \mathbb{E}(X_0) + \mathbb{E}(N_t) \cdot m$$

and thence for $m < \infty$

$$\frac{1}{m} - \frac{\mathbb{E}(X_0)}{m \cdot t} < \frac{\mathbb{E}(N_t)}{t}$$

for all $t \geq 0$. For $\mathbb{E}(X_0) < \infty$ and $t \to \infty$, this yields the bound

$$\liminf_{t\to\infty} \frac{\mathbb{E}(N_t)}{t} \geq \frac{1}{m}$$

which trivially holds for the case $m = \infty$.
Now it remains to show that $\limsup_{t\to\infty} \mathbb{E}(N_t)/t \leq 1/m$. To this aim we consider the truncated renewal process, denoted by $\tilde{\mathcal{N}}$, with the same delay $\tilde{X}_0 = X_0$ but renewal intervals $\tilde{X}_n = \min(X_n, M)$ for all $n \in \mathbb{N}$, with M being a fixed constant. Denote further $\tilde{m} = \mathbb{E}(\tilde{X}_1)$.
Because of $\tilde{X}_n \leq M$ the bound $\sum_{n=0}^{\tilde{N}_t} \tilde{X}_n \leq t + M$ holds almost certainly for all $t \geq 0$. Taking expectations and applying Wald's lemma, we obtain

$$\mathbb{E}(X_0) + \mathbb{E}(\tilde{N}_t) \cdot \tilde{m} = \mathbb{E}\left(\sum_{n=0}^{\tilde{N}_t} \tilde{X}_n\right) \leq t + M$$

For $\mathbb{E}(X_0) < \infty$ and $t \to \infty$, this yields

$$\limsup_{t\to\infty} \frac{\mathbb{E}(\tilde{N}_t)}{t} \leq \frac{1}{\tilde{m}}$$

Since $\tilde{X}_n \leq X_n$ for all $n \in \mathbb{N}$, we know that $\tilde{N}_t \geq N_t$ for all $t \geq 0$. Thus we obtain further

$$\limsup_{t\to\infty} \frac{\mathbb{E}(N_t)}{t} \leq \frac{1}{\tilde{m}}$$

for any constant M. Now the result follows for $M \to \infty$.
□

Remark 4.13 In view of theorem 4.4 one might be tempted to think that this trivially implied the statement of the above theorem 4.12. However, the following example shows that a limit with probability one in general does not imply a limit in expectation.
Let U denote a random variable which is uniformly distributed on the interval $]0, 1[$. Further define the random variables $(V_n : n \in \mathbb{N})$ by

$$V_n := \begin{cases} 0, & U > 1/n \\ n, & U \leq 1/n \end{cases}$$

Since $U > 0$ with probability one, we obtain the limit

$$V_n \to 0, \qquad n \to \infty$$

with probability one. On the other hand, the expectation for V_n is given by

$$\mathbb{E}(V_n) = n \cdot \mathbb{P}(U \leq 1/n) = n \cdot \frac{1}{n} = 1$$

for all $n \in \mathbb{N}$ and thus $\mathbb{E}(V_n) \to 1$ as $n \to \infty$.

4.4 Residual Life Times and Stationary Renewal Processes

Choose any time $t \geq 0$. Denote the duration from t until the next arrival by $B_t := S_{N_t+1} - t$ and call it the **residual life time** (or the **excess life**) at t. Further we define $A_t := t - S_{N_t}$ and call A_t the **age** at t. These definitions imply the equality $\{A_t > x\} = \{B_{t-x} > x\}$ for all $x < t$.
For a **stationary renewal process** we would postulate that the distribution of the counts in an interval $[s, s + t]$ be independent of s and thus equal the distribution of N_t. If this holds for a process $\mathcal{N}$, then we also say that $\mathcal{N}$ has **stationary**

increments. This implies in particular that the distribution of the residual life time must be independent of t, i.e. it coincides with the distribution of B_0 and hence of X_0. We first guess that it satisfies

$$\mathbb{P}(X_0 \leq x) = \frac{1}{m} \int_0^x (1 - F(y))\, dy \tag{4.6}$$

for all $x \geq 0$, where F denotes the distribution function of X_1 and further $m = \mathbb{E}(X_1) < \infty$. Indeed we can show

Theorem 4.14 *For a renewal process $\mathcal{N}$ defined by (4.6) the following properties hold:*
(1) $\mathbb{E}(N_t) = t/m$ *for all* $t \geq 0$
(2) $\mathbb{P}(B_t \leq x) = m^{-1} \int_0^x (1 - F(y))\, dy$ *for all* $t \geq 0$
(3) $\mathcal{N}$ has stationary increments.

Proof: (1) The distribution G of X_0 has a density $g(t) = \frac{1}{m}(1 - F(t))$ Hence the Laplace–Stieltjes transform (LST) of G is

$$\begin{aligned}\tilde{G}(s) &= \int_0^\infty e^{-st} \frac{1}{m}(1 - F(t))\, dt = \frac{1}{m} \left(\int_0^\infty e^{-st}\, dt - \int_0^\infty e^{-st} F(t)\, dt \right) \\ &= \frac{1}{m} \left(\frac{1}{s} - \frac{1}{s} \int_0^\infty e^{-st}\, dF(t) \right) = \frac{1 - \tilde{F}(s)}{sm}\end{aligned}$$

with $\tilde{F}(s)$ denoting the LST of F. According to (4.1) we have the representation $\mathbb{E}(N_t) = G * \sum_{n=0}^\infty F^{*n}(t)$ for all $t \geq 0$. Hence the LST of $M(t) := \mathbb{E}(N_t)$ is given by

$$\tilde{M}(s) = \frac{\tilde{G}(s)}{1 - \tilde{F}(s)} = \frac{1}{sm}$$

for all $s > 0$, and thus coincides with the LST of the measure dx/m. Since the LST uniquely determines a function on $[0, \infty[$, this proves the first statement.
(2) The joint distributions

$$\begin{aligned}\mathbb{P}(B_t > x, N_t = 0) &= 1 - G(t + x) \\ \mathbb{P}(B_t > x, N_t = n) &= \int_0^\infty \mathbb{P}(B_t > x, N_t = n | S_n = y)\, dG * F^{*n-1}(y) \\ &= \int_0^t (1 - F(t + x - y))\, dG * F^{*n-1}(y)\end{aligned}$$

for $n \geq 1$ are immediate from the definition. Abbreviating $F^c(x) := 1 - F(x)$, $G^c(x) := 1 - G(x)$, and denoting $M(t) := \mathbb{E}(N_t)$, we can write

$$\begin{aligned}
\mathbb{P}(B_t > x) &= \sum_{n=0}^{\infty} \mathbb{P}(B_t > x, N_t = n) \\
&= G^c(t+x) + \sum_{n=1}^{\infty} \int_0^t F^c(t+x-y)\, dG * F^{*n-1}(y) \\
&= G^c(t+x) + \int_0^t F^c(t+x-y)\, d\left(\sum_{n=1}^{\infty} G * F^{*n-1}\right)(y) \\
&= G^c(t+x) + \int_0^t F^c(t+x-y)\, dM(y)
\end{aligned}$$

Using statement (1) and the definition of G, we obtain

$$\begin{aligned}
\mathbb{P}(B_t > x) &= 1 - \frac{1}{m}\int_0^{t+x} (1 - F(y))\, dy + \frac{1}{m}\int_0^t (1 - F(t+x-y))\, dy \\
&= 1 - \frac{1}{m}\int_0^x (1 - F(y))\, dy
\end{aligned}$$

which proves the second statement.
(3) The difference $N_{t+s} - N_s$ simply counts the number N'_t of events in time t of the renewal process $\mathcal{N}'$ with the same distribution F of X_1 but a delay $X'_0 \sim B_s$. Now statement (2) shows that $X_0 \sim B_s = B_0$. Hence we obtain $N'_t = N_t = N_{t+s} - N_s$ in distribution, which was to be proven.
□
Because of the results above a renewal process which satisfies condition (4.6) is called **stationary renewal process**. As one would expect, also the mean residual life time $\mathbb{E}(B_t)$ of a stationary renewal process coincides with the limit of the mean residual life time of an ordinary renewal process:

Lemma 4.15 *For a non–negative random variable X the nth moment can be expressed by*

$$\mathbb{E}(X^n) = \int_0^{\infty} \mathbb{P}(X > x) \cdot n x^{n-1}\, dx$$

Proof: This follows simply by writing

$$\mathbb{E}(X^n) = \int_0^{\infty} \mathbb{P}(X^n > z)\, dz = \int_0^{\infty} \mathbb{P}(X > \sqrt[n]{z})\, dz$$

and substituting $x = \sqrt[n]{z}$ with $nx^{n-1}\,dx = dz$.
□

Theorem 4.16 *For a stationary renewal process with $\mathbb{E}(X_1^2) < \infty$ the mean residual life time is given by*

$$\mathbb{E}(B_t) = \frac{\mathbb{E}(X_1^2)}{2m}$$

independently of $t \geq 0$.

Proof: Using part (2) of theorem 4.14, we obtain

$$\begin{aligned}\mathbb{E}(B_t) &= \int_0^\infty \mathbb{P}(B_t > y)\,dy = \frac{1}{m}\int_{y=0}^\infty \int_{x=y}^\infty (1 - F(x))\,dx\,dy \\ &= \frac{1}{m}\int_{x=0}^\infty \int_{y=0}^x (1 - F(x))\,dy\,dx = \frac{1}{m}\int_{x=0}^\infty \mathbb{P}(X_1 > x)\cdot x\,dx\end{aligned}$$

and the statement follows from lemma 4.15.
□

Example 4.17 Waiting time at a bus stop
Consider a bus stop where buses are scheduled to arrive in intervals of length T. However, due to traffic variations the real inter–arrival times are uniformly distributed within intervals $[T - a, T + a]$ with some $a > 0$. Now suppose that somebody arrives at the bus stop "at random", i.e. without knowing the bus schedule. Then we can model the mean waiting time for the next bus by the mean residual life time $\mathbb{E}(B_t)$ in a stationary renewal process with distribution $X_1 \sim U(T - a, T + a)$. We obtain

$$\mathbb{E}(X_1^2) = \frac{1}{2a}\int_{T-a}^{T+a} x^2\,dx = \frac{1}{2a}\cdot\frac{1}{3}\left(6T^2a + 2a^3\right) = T^2 + \frac{a^2}{3}$$

and by theorem 4.16

$$\mathbb{E}(B_t) = \frac{T^2 + \frac{a^2}{3}}{2\cdot T} = \frac{T}{2} + \frac{a^2}{6\cdot T}$$

Thus the mean waiting time for random inter–arrival times (meaning $a > 0$) is longer than it would be for deterministic ones (namely $T/2$). This phenomenon is called the **waiting time paradox**.

Notes

A classical presentation of renewal theory is chapter 11 in Feller [6]. The presentation in this chapter is largely adapted to Ross [18, 19] as well as Karlin and Taylor [11].

Exercise 4.18 Prove theorem 4.3.

Exercise 4.19 In the proof for Wald's lemma 4.11 we have used the relation $\mathbb{E}(S) = \sum_{n=0}^{\infty} \mathbb{P}(S > n)$. For a positive continuous distribution function F, the equivalent is $\mathbb{E}(F) = \int_0^{\infty} (1 - F(y))\, dy$. Give a proof for these equations.

Exercise 4.20 Show for a Poisson process $\mathcal{N}$ with intensity $\lambda > 0$ that

$$\mathbb{P}(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

for all $t \geq 0$ and $k \in \mathbb{N}_0$, and $\mathbb{E}(N_t) = \lambda \cdot t$.

Exercise 4.21 A plumber receives orders at time intervals which are distributed exponentially with parameter λ. As soon as he has received an order he goes to work, which takes an exponentially distributed time with parameter μ. During work he cannot receive any orders. Assume that at time zero the plumber is working. Give a model of the plumber receiving orders in terms of a renewal process and determine the density of the renewal intervals' distribution.

Exercise 4.22 An intelligence agency eavesdrops on telephone calls automatically. If there occurs a suspicious sequence of words, a closer investigation is initiated. The probabilitiy for such a sequence is one in a thousand for every call. The length of a call is distributed exponentially with a mean of 20 seconds. How long is the expected amount of time before a closer investigation begins? Use Wald's lemma.

Exercise 4.23 Let $\mathcal{N} = (N_t : t \geq 0)$ denote an ordinary renewal process with $X_1 \sim F$. Show that the current life time X_{N_t} satisfies

$$\mathbb{P}(X_{N_t} > x) \geq 1 - F(x)$$

for all $x \geq 0$.

Exercise 4.24 Show that the age A_t of a stationary renewal process is distributed as

$$\mathbb{P}(A_t \leq x) = \frac{1}{m} \int_0^x (1 - F(y))\, dy$$

independently of $t \geq 0$.

Chapter 5

Appendix

5.1 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

$$\int_C X_0 dP = \int_C X dP \tag{5.1}$$

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 (5.1). Typical special cases and examples are

Example 5.1 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 5.2 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

$$\mathbb{E}(\mathbb{E}(X|\mathcal{C}_2)|\mathcal{C}_1) = \mathbb{E}(\mathbb{E}(X|\mathcal{C}_1)|\mathcal{C}_2) = \mathbb{E}(X|\mathcal{C}_1) \tag{5.2}$$

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

$$\mathbb{E}(XY|\mathcal{C}) = X \cdot \mathbb{E}(Y|\mathcal{C}) \tag{5.3}$$

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 $y \in \Omega'$.

5.2 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.

5.2.1 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 [7], section II.4.

Theorem 5.3 *There is a probability measure $\mathbb{P}$ on $(S^\infty, \mathcal{B}^\infty)$ satisfying*

$$\mathbb{P}(C_{0,\ldots,k}(A_0 \times \ldots \times A_k)) = \int_{A_0} d\mu(x_0) \int_{A_1} P_1(x_0, dx_1) \ldots \\ \ldots \int_{A_{k-1}} P_{k-1}(x_{k-2}, dx_{k-1})\, P_k(x_{k-1}, A_k) \quad (5.4)$$

for all $k \in \mathbb{N}_0$, $A_0, \ldots, A_k \in \mathcal{B}$. The measure $\mathbb{P}$ is uniquely determined by the system (5.4) 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 $(S^\infty, \mathcal{B}^\infty)$.

5.2.2 Stochastic processes

Let S denote a Polish (i.e. a complete separable metric) space, and $\mathcal{B}$ the Borel σ–algebra on S. Define Ω as the set of all functions $f : \mathbb{R}_0^+ \to S$. In order to construct an appropriate σ–algebra on Ω, we again start from the cylinder sets

$$C_{t_1,\ldots,t_k}(A) = \{f \in \Omega : (f(t_1), \ldots, f(t_k)) \in A\}$$

for $k \in \mathbb{N}$, $t_1 < \ldots < t_k \in \mathbb{R}_0^+$, and $A \in \mathcal{B}^k$. Denote the set of all cylinders in Ω by $\mathcal{C}$. Again, $\mathcal{C}$ forms an algebra of sets and we can define $\mathcal{A} := \sigma(\mathcal{C})$ as the minimal σ–algebra containing $\mathcal{C}$.
Let $\mathcal{M} = \{\mu_{t_1,\ldots,t_k} : k \in \mathbb{N}, t_1, \ldots, t_k \in \mathbb{R}_0^+\}$ denote a family of probability distributions with
(C1) For all $k \in \mathbb{N}$, $t_1, \ldots, t_k \in \mathbb{R}_0^+$, and $A \in \mathcal{B}^k$

$$\mu_{t_1,\ldots,t_k,t_{k+1}}(A \times S) = \mu_{t_1,\ldots,t_k}(A)$$

(C2) For all $k \in \mathbb{N}$ and permutations $\pi : \{1, \ldots, k\} \to \{1, \ldots, k\}$

$$\mu_{\pi(t_1,\ldots,t_k)}(\pi(A)) = \mu_{t_1,\ldots,t_k}(A)$$

Then the family $\mathcal{M}$ is called **compatible**.

Remark 5.4 Condition (C1) ensures that the distributions are consistent with each other, while condition (C2) is merely notational.

The following extension theorem by Kolmogorov is proven in Gikhman and Skorokhod [9], section 3.2.

Theorem 5.5 *Let $\{\mu_{t_1,\ldots,t_k} : k \in \mathbb{N}, t_1, \ldots, t_k \in \mathbb{R}_0^+\}$ denote a compatible family of probability measures. Then there is a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$ with*

$$\mathbb{P}(\{f \in \Omega : (f(t_1), \ldots, f(t_k)) \in A\}) = \mu_{t_1,\ldots,t_k}(A) \tag{5.5}$$

for all $k \in \mathbb{N}$, $t_1, \ldots, t_k \in \mathbb{R}_0^+$, and $A \in \mathcal{B}^k$. The measure $\mathbb{P}$ is uniquely determined by the system (5.5) of equations.

Based on this, we define a **stochastic process** with Polish state space S as a family $X = (X_t : t \in \mathbb{R}_0^+)$ of S–valued random variables which are distributed according to a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{A})$. An element $\omega \in \Omega$ is an arbitrary function $\omega : \mathbb{R}_0^+ \to S$. It is also called a **path** of X. If we want to state that the support of $\mathbb{P}$ consists of a special class of functions (say right–continuous ones), then we say that X is a stochastic process with right–continuous paths. The above family $\mathcal{M}$ of probability measures is called the set of **finite–dimensional marginal distributions** for X.
Due to theorem 5.5 a Markov process is uniquely defined by its initial distribution and the family of transition probabilities, since they determine all finite–dimensional marginal distributions. Further our constructions of Markov processes, renewal processes, and semi–Markov processes yield compatible sets of finite–dimensional marginal distributions, hence by theorem 5.5 a probability measure $\mathbb{P}$ for the respective process.

5.3 Transforms

In several parts of the present book, it is essential to argue via transforms of distributions. The necessary background for these shall be presented shortly in this section. For discrete distributions on $\mathbb{N}_0$ we will introduce z–transforms, while for distributions on $\mathbb{R}_0^+$ the Laplace–Stieltjes transform will be useful.

5.3.1 z–transforms

Let X denote a $\mathbb{N}_0$–valued random variable with distribution $A = (a_n : n \in \mathbb{N}_0)$, i.e. $\mathbb{P}(X = n) = a_n$ for all $n \in \mathbb{N}_0$. Then the power series

$$A^*(z) := \sum_{n=0}^{\infty} a_n z^n \tag{5.6}$$

converges absolutely for $z \in \mathbb{C}$ with $|z| \leq 1$ and is analytic in this region. We note that $A^*(z) = \mathbb{E}(z^X)$. If $A(z)$ is a given power series for a distribution $(a_n : n \in \mathbb{N}_0)$, then the probabilities a_n can be derived as

$$a_n = \frac{1}{n!} \frac{d^n}{dz^n} A(z) \Big|_{z=0}$$

for all $n \in \mathbb{N}_0$. Thus the mapping between discrete distributions on $\mathbb{N}_0$ and the power series in (5.6) is bijective, and we may call $A^*(z)$ the (uniquely determined) **z–transform** of X (also: of the distribution A).

Example 5.6 For a Dirac distribution on $k \in \mathbb{N}_0$ with

$$a_n = \begin{cases} 1, & n = k \\ 0, & n \neq k \end{cases}$$

we obtain $A^*(z) = z^k$.

Example 5.7 Let A denote the geometric distribution with some parameter $p \in]0, 1[$, i.e.

$$a_n = (1-p)p^n$$

for all $n \in \mathbb{N}_0$. The z–transform of A is given by

$$A^*(z) = (1-p)\sum_{n=0}^{\infty} p^n z^n = \frac{1-p}{1-pz}$$

for all $|z| \leq 1$.

A very useful feature is the behaviour of the z–transform with respect to the convolution of two distributions. Let $A = (a_n : n \in \mathbb{N}_0)$ and $B = (b_n : n \in \mathbb{N}_0)$ denote two distributions on $\mathbb{N}_0$. The convolution $C = A * B$ of A and B is defined as the distribution $C = (c_n : n \in \mathbb{N}_0)$ with

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}$$

for all $n \in \mathbb{N}_0$. For the z–transform of C we obtain

$$\begin{aligned} C^*(z) &= \sum_{n=0}^{\infty} c_n z^n = \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_k b_{n-k} z^n = \sum_{n=0}^{\infty} a_k z^k \sum_{n=k}^{\infty} b_{n-k} z^{n-k} \\ &= A^*(z) \cdot B^*(z) \end{aligned}$$

for all $|z| \leq 1$.
This means that the z–transform of a convolution $A * B$ equals the product $A^*(z) \cdot B^*(z)$ of the z–transform of A and B. In terms of random variables we have

the following representation: Let X and Y denote two independent $\mathbb{N}_0$–valued random variables. Then the z–transform of the sum $X + Y$ equals the product of the z–transforms of X and Y, i.e.

$$\mathbb{E}\left(z^{X+Y}\right) = \mathbb{E}\left(z^X\right) \cdot \mathbb{E}\left(z^Y\right)$$

for all $|z| \leq 1$.

5.3.2 Laplace–Stieltjes transforms

Let X denote an $\mathbb{R}_0^+$–valued random variable with distribution function F. The **Laplace–Stieltjes transform** (LST) of X (or: of F) is defined by

$$F^*(s) := \int_0^\infty e^{-st} dF(t) = \mathbb{E}\left(e^{-sX}\right)$$

for all $s \in \mathbb{C}$ with $Re(s) \geq 0$. The LST uniquely determines its underlying distribution.

Example 5.8 Let X be exponentially distributed with parameter λ, i.e. X has the distribution function $F(t) = 1 - e^{-\lambda t}$ with Lebesgue density $f(t) = \lambda e^{-\lambda t}$. Then

$$F^*(s) = \int_0^\infty e^{-st} \lambda e^{-\lambda t}\, dt = \frac{\lambda}{s + \lambda}$$

for $Re(s) \geq 0$.

Example 5.9 For the Dirac distribution δ_x on $x \in \mathbb{R}_0^+$ we obtain

$$\delta_x^*(s) = \int_0^\infty e^{-st} dF(t) \qquad \text{with} \qquad F(t) = \begin{cases} 0, & t < x \\ 1, & t \geq x \end{cases}$$

and hence

$$\delta_x^*(s) = e^{-sx}$$

for $Re(s) \geq 0$.

Like the z–transform, the LST is very useful for dealing with convolutions. Let X and Y denote two independent $\mathbb{R}_0^+$–valued random variables. Then the LST of the sum $X + Y$ equals the product of the LSTs of X and Y, i.e.

$$\mathbb{E}\left(e^{-s(X+Y)}\right) = \mathbb{E}\left(e^{-sX}\right) \cdot \mathbb{E}\left(e^{-sY}\right)$$

for all $s \in \mathbb{C}$ with $Re(s) \geq 0$.

Notes

For more on z–transforms see e.g. Juri [10], or the collection of results in Kleinrock [12], appendix I. For Laplace–Stieltjes transforms see chapter XIII in Feller [6] or again Kleinrock [12], appendix I.

5.4 Gershgorin's circle theorem

An important theorem to find bounds for the eigenvalues of a matrix has been developed by Gershgorin in 1938. For ease of reference it shall be presented in this section. Let $A = (a_{ij})_{i,j \le m}$ denote a square matrix of dimension m with entries $a_{ij} \in \mathbb{C}$. The following theorem is called **Gershgorin's circle theorem**.

Theorem 5.10 *All eigenvalues of the matrix A lie in the union $C := \bigcup_{i=1}^{m} C_i$ of the circles*

$$C_i = \left\{ z \in \mathbb{C} : |z - a_{ii}| \le \sum_{k \neq i} |a_{ik}| \right\}$$

Proof: Let $x^{(\nu)}$ denote an eigenvector to the eigenvalue λ_ν of A, i.e. $Ax^{(\nu)} = \lambda_\nu x^{(\nu)}$. This implies

$$\sum_{k=1}^{m} a_{ik} x_k^{(\nu)} = \lambda_\nu x_i^{(\nu)} \tag{5.7}$$

for all $i \le m$. Since an eigenvector is determined only up to a scalar multiplicative, we can assume without loss of generality that there is a component

$$x_{i_0}^{(\nu)} = \max_{1 \le j \le m} \left| x_j^{(\nu)} \right| = 1$$

of the vector $x^{(\nu)}$. Now (5.7) yields for $i = i_0$ the relation

$$\sum_{k \neq i_0} a_{i_0,k} x_k^{(\nu)} = (\lambda_\nu - a_{i_0,i_0})\, x_{i_0}^{(\nu)} = \lambda_\nu - a_{i_0,i_0}$$

which implies by the triangle inequality

$$|\lambda_\nu - a_{i_0,i_0}| \le \sum_{k \neq i_0} |a_{i_0,k}| \cdot \left| x_k^{(\nu)} \right| \le \sum_{k \neq i_0} |a_{i_0,k}|$$

Since every eigenvalue satisfies at least one such inequality, the proof is complete. □

Corollary 5.11 *If A is diagonally dominated, i.e. if*

$$|a_{ii}| > \sum_{k \neq i} |a_{ik}|$$

holds for all $1 \leq i \leq m$, then the matrix A is invertible.

Proof: The strict inequality of the assumption implies that $a_{ii} \neq 0$ for all $i \leq m$. Applying theorem 5.10 yields a restriction

$$|\lambda| \geq |a_{ii}| - |a_{ii} - \lambda| \geq |a_{ii}| - \sum_{k \neq i} |a_{ik}| > 0$$

for every eigenvalue λ of A. Therefore the matrix A has no eigenvalue zero and thus is invertible.
□

Bibliography

[1] S. Asmussen, O. Nerman, and M. Olsson. Fitting phase-type distributions via the EM algorithm. *Scand. J. Stat.*, 23(4):419–441, 1996.

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

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

[4] K. Chung. *Markov chains with stationary transition probabilities*. Springer-Verlag, Berlin etc., 1960.

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

[6] W. Feller. *An introduction to probability theory and its applications. Vol. II. 2nd ed.* New York etc.: John Wiley and Sons, 1971.

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

[8] I. Gihman and A. Skorohod. *The theory of stochastic processes II*. Berlin etc.: Springer-Verlag, 1975.

[9] I. Gikhman and A. Skorokhod. *Introduction to the theory of random processes*. Saunders, 1969.

[10] E. Juri. *Theory and Application of the z–Transform Method*. Wiley, New York, 1964.

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

[12] L. Kleinrock. *Queueing systems. Vol. I: Theory.* New York etc.: John Wiley & Sons, 1975.

[13] Y. Lam. A Note on the Optimal Replacement Problem. *Adv. Appl. Prob.*, 20:479–482, 1988.

[14] G. Latouche and V. Ramaswami. *Introduction to matrix analytic methods in stochastic modeling.* Philadelphia, PA: SIAM, 1999.

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

[16] M. F. Neuts. Probability distributions of phase type. In *Liber Amicorum Prof. Emeritus H. Florin*, pages 173–206. Department of Mathematics, University of Louvain, Belgium, 1975.

[17] M. F. Neuts. *Matrix–Geometric Solutions in Stochastic Models*. Johns Hopkins University Press, Baltimore, 1981.

[18] S. Ross. *Applied Probability Models with Optimization Applications*. Holden–Day, San Francisco, 1970.

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

[20] D. Shi, J. Guo, and L. Liu. SPH–Distributions and the Rectangle–Iterative Algorithm. In S. R. Chakravarthy and A. S. Alfa, editors, *Matrix–Analytic Methods in Stochastic Models*, pages 207–224, New York etc., 1997. Marcel Dekker.

[21] D. Shi and L. Liu. Markovian Models for Non–negative Random Variables. In A. S. Alfa and S. R. Chakravarthy, editors, *Advances in Matrix Analytic Methods for Stochastic Models*, pages 403–427. Notable Publications, 1998.

[22] W. Smith. Regenerative Stochastic Processes. *Proceedings Royal Society, Series A*, 232:6–31, 1955.

[23] W. Smith. Renewal Theory and Its Ramifications. *Journal of the Royal Statistical Society, Series B*, 20:243–302, 1958.