FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

Department of Probability and Mathematical Statistics

Stochastic Processes 1

Lecture Notes, December 2018

Zuzana Prášková

Contents

Chapter 1

Introduction

1.1 Definition and basic properties of a stochastic process

Definition 1.1. Let $(\Omega, \mathcal{A}, P)$ be a probability space, $T \subset \mathbb{R}$. A family of random variables $\{X_t, t \in T\}$ defined on $(\Omega, \mathcal{A}, P)$ is called to be *stochastic (random) process.*

If $T = \mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ or $T = \mathbb{N}_0 = \{0, 1, \dots\}$,

$\{X_t, t \in T\}$ is called *discrete time stochastic process* or *time series.* If $T = [a, b]$, where $-\infty \leq a < b \leq \infty$, $\{X_t, t \in T\}$ is said to be *continuous time process.*

The pair $(S, \mathcal{E})$, where S is a set of values of random variables X_t and $\mathcal{E}$ is $\sigma-$algebra of subsets of S, is called *state space* of the process $\{X_t, t \in T\}$. If X_t are discrete-valued only, we speak about the *discrete state process*, otherwise, if X_t take values from an interval, we speak on *continuous state process.* If $S = \mathbb{R}$ and $\mathcal{E}$ is a Borel $\sigma-$algebra, $\{X_t, t \in T\}$ is a real-valued process.

A stochastic process $\{X_t, t \in T\}$ can be considered to be a function of two variables ω, t. Given $t \in T$ fixed, $X_t = X_t(.)$ is a random variable defined on Ω; for a fixed $\omega \in \Omega$ $X_{(.)} = X_{(.)}(\omega)$ is a real-valued function of t, which is called *trajectory of the process* $\{X_t, t \in T\}$.

For every finite subset $\{t_1, \dots, t_n\} \subset T$ there exists a system of random variables $X_{t_1}, \dots, X_{t_n}$, with joint distribution function

$$F_{t_1,\dots,t_n}(x_1, \dots, x_n) = P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n).$$

For $n \in \mathbb{N}$ and $t_1, \dots, t_n \in T$, the system of distribution functions $\{F_{t_1,\dots,t_n}(x_1, \dots, x_n)\}$ has the following properties:

a) for any permutation $i_1, \dots, i_n$ of $1, \dots, n$,

$$F_{t_{i_1},\dots,t_{i_n}}(x_{i_1}, \dots, x_{i_n}) = F_{t_1,\dots,t_n}(x_1, \dots, x_n),$$

b)

$$\lim_{x_{n+1}\to\infty} F_{t_1,\dots,t_n,t_{n+1}}(x_1, \dots, x_n, x_{n+1}) = F_{t_1,\dots,t_n}(x_1, \dots, x_n).$$

A system of distribution functions that satisfies *a*) a *b*) is called to be *consistent.* For any stochastic process there exists a consistent system of distribution functions. On the other hand, the following theorem can be proved.

Theorem 1.1. (Daniell-Kolmogorov) *Let $\{F_{t_1,\dots,t_n}(x_1, \dots, x_n)\}$ be a consistent system of distribution functions. Then there exists a stochastic process $\{X_t, t \in T\}$ such that, for every $n \in \mathbb{N}$, any $t_1, \dots, t_n \in T$ and any real numbers $x_1, \dots, x_n$*

$$P(X_{t_1} \le x_1, \dots, X_{t_n} \le x_n) = F_{t_1,\dots,t_n}(x_1, \dots, x_n).$$

Proof. Štěpán (1987), věta (Theorem) I.10.3. □

The distribution of a stochastic process $\{X_t, t \in T\}$ is uniquely determined by the distribution of all finite-dimensional random vectors $(X_{t_1}, \dots, X_{t_n})$.

Definition 1.2. Let $\{X_t, t \in T\}$ be a stochastic process such that for every $t \in T$ EX_t exists. Then the function $\mu_t = EX_t$ defined on T is called *mean value of the process* $\{X_t, \in T\}$. If $E|X_t|^2 < \infty$ for all $t \in T$, the function of two variables on $T \times T$ defined by $R(s,t) = E(X_s - \mu_s)(X_t - \mu_t)$ is called *autocovariance function of the process* $\{X_t, \in T\}$. The value $R(t,t)$ is called *variance of the process* at time t.

Definition 1.3. We say that a stochastic process $\{X_t, t \in T\}$ is *strictly stationary,* if for every $n \in \mathbb{N}$, for any real-valued $x_1, \dots, x_n$ and for any $t_1, \dots, t_n$ and h such that $t_k \in T, t_k + h \in T, 1 \le k \le n$,

$$F_{t_1,\dots,t_n}(x_1, \dots, x_n) = F_{t_1+h,\dots,t_n+h}(x_1, \dots, x_n).$$

It holds from the definition of strict stationarity that all the random variables X_t are identically distributed.

Remark. A finite second order moments stochastic process is *weakly stationary* or *second order stationary* if its mean value is constant and its autocovariance function $R(s,t)$ is a function of difference $t - s$, *only*.

Definition 1.4. Let $\{X_t, t \in T\}, \{Y_t, t \in T\}$ be stochastic processes defined on the same probability space $(\Omega, \mathcal{A}, P)$ with values in the same state space $(S, \mathcal{E})$. We say that processes $\{X_t\}, \{Y_t\}$ are *stochastically equivalent* if for every $t \in T$

$$P(X_t = Y_t) = P(\{\omega : X_t(\omega) = Y_t(\omega)\}) = 1.$$

If $\{X_t\}$ is a stochastic process on $(\Omega, \mathcal{A}, P)$ and $\{Y_t\}$ is stochastically equivalent to $\{X_t\}$, we say that $\{Y_t\}$ is *a stochastic version of* the process $\{X_t\}$.

Stochastically equivalent processes have the same all finite-dimensional distributions: for every $n \in \mathbb{N}, t_1, \ldots, t_n \in T$ and any real-valued $x_1, \ldots, x_n$ we have

$$\begin{aligned} P(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n) &= P(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n, X_{t_1} = Y_{t_1}, \ldots X_{t_n} = Y_{t_n}) \\ &= P(Y_{t_1} \le x_1, \ldots, Y_{t_n} \le x_n). \end{aligned}$$

Remark. Stochastically equivalent processes do not need to have the same trajectories. Let, e.g., $\Omega = [0, 1], \mathcal{A}$ be the $\sigma-$algebra of Borel subsets of $[0, 1]$ and P be the Lebesgue measure on $[0, 1]$, let $T = [0, 1]$. Consider processes $\{X_t, t \in T\}$ and $\{Y_t, t \in T\}$ defined on $(\Omega, \mathcal{A}, P)$ by

$$X_t(\omega) = 0, \quad \omega \in \Omega, \ t \in T,$$

$$Y_t(\omega) = \begin{cases} 0, & t \ne \omega, \\ 1, & t = \omega. \end{cases}$$

Then for every $t \in T$

$$P(\{\omega : X_t(\omega) = Y_t(\omega)\}) = P(\{\omega : \omega \ne t\}) = 1,$$

which means that $\{X_t\}, \{Y_t\}$ are stochastically equivalent, but they have different trajectories.

Definition 1.5. A stochastic process $\{X_t, t \in T\}$ is called to be *stochastically continuous (continuous in probability)* at point $t_0 \in T$, if for every $\varepsilon > 0$

$$\lim_{t \to t_0} P(|X_t - X_{t_0}| > \varepsilon) = 0.$$

The process is stochastically continuous if it is stochastically continuous at every point of T.

Remark. A stochastically continuous process need not to have continuous trajectories. Let us consider, e.g., a sequence of independent and identically distributed random variables $\{T_k, k \geq 1\}$ with a continuous distribution function $F(t) = P(T_k \leq t)$ and define the process $\{X_t, t \geq 0\}$ by

$$X_t = \sum_{k=1}^{n} I(T_k \leq t), \quad t \geq 0,$$

where $I(A)$ is the indicator of an event A and $n \in \mathbb{N}$. Obviously, $0 \leq X_{t_1} \leq X_{t_2}$ for every $0 \leq t_1 < t_2$. The trajectories of the process $\{X_t, t \geq 0\}$ are non-decreasing step functions with jumps at points $T_{(1)} < \cdots < T_{(n)}$, where $T_{(1)}, \ldots T_{(n)}$ are order statistics of $T_1, \ldots, T_n$. However, the process $\{X_t, t \geq 0\}$ is stochastically continuous since, according to the Markov inequality, for every $h > 0$ and $\varepsilon > 0$

$$P(|X_{t+h} - X_t| > \varepsilon) \leq \frac{1}{\varepsilon} E|X_{t+h} - X_t| = \frac{n}{\varepsilon} EI(t < T_1 \leq t + h) = \frac{n}{\varepsilon}(F(t+h) - F(t))$$

and due to the continuity of the distribution function F, the last expression converges to 0 as $h \to 0_+$; similarly, $P(|X_{t-h} - X_t| > \varepsilon) \to 0$ as $h \to 0_+$.

Let $\{X_t, t \in T\}$ be a stochastic process defined on $(\Omega, \mathcal{A}, P)$. The mapping $X_t(.) : \Omega \to \mathbb{R}$ is $\mathcal{A}-$measurable for every $t \in T$, i.e., $\{\omega : X_t(\omega) \in B\} \in \mathcal{A}$ for every Borel subset $B \subset \mathbb{R}$. If T is not countable, sets

$$\{\omega : X_t(\omega) \in B, t \in T\} = \bigcap_{t \in T} \{\omega : X_t(\omega) \in B\}$$

generally do not belong to $\mathcal{A}$; then for a continuous time stochastic process $\{X_t, t \in T\}$, functionals like $\sup_{t \in T} X_t$, $\inf_{t \in T} X_t$ need not be random variables. This founding leads us to the following definition.

Definition 1.6. A stochastic process $\{X_t, t \in T\}$, where $T \subset \mathbb{R}$ is an interval, is called to be *separable,* if there exist a countable dense subset $D \subset T$ and an event $\Lambda \subset \Omega$ of zero probability such that

$$\{\omega : X_t(\omega) \in C, t \in J \cap D\} \setminus \{\omega : X_t(\omega) \in C, t \in J \cap T\} \subset \Lambda$$

for any closed set $C \subset \mathbb{R}$ and any open interval $J \subset T$. The countable set D is said to be *separant* of the process $\{X_t\}$.

Since

$$\{\omega : X_t(\omega) \in C, t \in J \cap D\} \supset \{\omega : X_t(\omega) \in C, t \in J \cap T\},$$

it follows from the last definition that

$$\Lambda^c \cap \{\omega : X_t(\omega) \in C, t \in J \cap D\} = \Lambda^c \cap \{\omega : X_t(\omega) \in C, t \in J \cap T\},$$

and since $\Lambda^c \in \mathcal{A}$ and the left-hand side is an event in $\mathcal{A}$, so is the right-hand side.

Remark. If $\{X_t, t \in T\}$ is a separable process and Λ, D are as in Definition 1.6, then it can be shown that for every $\omega \in \Lambda^c, t \in T$ there is a sequence $\{t_n, n \in \mathbb{N}\} \subset D$ and $t_n \to t$ as $n \to \infty$ such that

$$X_t(\omega) = \lim_{n\to\infty} X_{t_n}(\omega) \ .$$

(Štěpán, 1987, I.10).

Definition 1.7. A stochastic process $\{X_t, t \in T\}$ is called *measurable*, if the mapping $(\omega, t) \to X_t(\omega)$ is $\mathcal{A} \otimes \mathcal{B}_T-$ measurable, where $\mathcal{B}_T$ is the $\sigma-$algebra of Borel subsets of T and $\mathcal{A} \otimes \mathcal{B}_T$ denotes the product $\sigma-$algebra.

Theorem 1.2. *Let $\{X_t, t \in T\}$ be a stochastically continuous process and $T \subset \mathbb{R}$ be an interval. Then there exists a version of the process (with states in $\overline{\mathbb{R}}$), that is separable and measurable.*

Proof. Štěpán (1987), 1.10.14. □

1.2 Examples of stochastic processes

Example 1.1. *White noise* is a process $\{X_t, t \in \mathbb{Z}\}$ of uncorrelated random variables with zero mean and the same finite variance. If X_t are independent and identically distributed we speak about the strict white noise.

Example 1.2. *Random walk on the real line.* Let $Y_1, Y_2, \dots$ be independent and identically distributed random variables taking values ± 1 with probability $1/2$. Define $X_0 = 0$ and for $n \in \mathbb{N}$ put $X_n = \sum_{j=1}^{n} Y_j$. The random variable X_n represents the position of a particle after n steps if we assume that it moves randomly on the integer-valued lattice (in $\mathbb{Z}$) with the same probability to the left and to the right.

The sequence $\{X_n, n \in \mathbb{N}_0\}$ is called random walk. Generally, any sum of independent identically distributed random variables is said to be a random walk.

Example 1.3. *Galton-Watson branching process* is a random sequence $\{X_n, n \in \mathbb{N}_0\}$, that represents numbers of individuals in generations $n = 0, 1 \ldots$. It is assumed that any individual of the $n-$the generation, $n \geq 0$, give a birth of a random number of offsprings in the generation $n+1$. The numbers of offsprings are equally distributed and independent both mutually and of the previous generation. Assuming that at the beginning $(n = 0)$ there is only one individual, the number of individuals in the $n-$th generation can be expressed by $X_n = U_{n1} + \cdots + U_{nX_{n-1}}$, where $U_{n1}, U_{n2}, \ldots$ are independent random variables with the same distribution as X_1, and independent of $X_{n-1}, X_{n-2}, \ldots, X_0$.

Example 1.4. *Poisson process.* Suppose that we register occurrences of some events during a time interval $[0, t]$, e.g., the number of particles registered in Geiger-Müller counter, the number of calls registered in a call center, or the number of insurance claims that register an insurance company. Assume that in the interval $(t, t+h]$, independently of t, exactly one event occurs with probability $\lambda h + o(h)$, more than one event with probability $o(h)$, and numbers of events that occur at disjoint time intervals are mutually independent random variables. The symbol $o(h)$ means that $o(h)/h \to 0$ as $h \to 0_+$ and λ is a positive constant

Let N_t denotes the number of events that occur in interval $[0, t]$. Then $\{N_t, t \geq 0\}$ is a stochastic process. Under the above assumptions, the random variables N_t have the Poisson distribution with the parameter λt,

$$P(N_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \quad k = 0, 1, \ldots$$

Stochastic process $\{N_t, t \geq 0\}$ is called the Poisson process and the constant λ is the *intensity of the Poisson process.*

Example 1.5. *The Wiener process (Brownian motion)* is a stochastic process $\{W_t, t \geq 0\}$, with the properties

(1) $W_0 = 0$ and the trajectories of $\{W_t, t \geq 0\}$ are continuous (real-valued) functions.

(2) For any $0 \leq t_1 < t_2 < \cdots < t_n$, the increments $W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent random variables.

(3) For any $0 \leq t < s$, the increments $W_s - W_t$ have normal distribution with zero mean and the variance $\sigma^2(s - t)$, where σ^2 is a positive constant.

The Wiener process is historically connected with the process of the Brownian motion, i. e., the random motion of particles suspended in a liquid resulting from their collision with fast-moving molecules of the liquid. The Wiener process plays an important role e.g. in physics, applied mathematics, probability and financial mathematics.

Chapter 2

Discrete time Markov chains

2.1 Basic properties

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\{X_n, n \in \mathbb{N}_0\}$, be a sequence of integer-valued random variables defined on it. Let S be a set of integers i such that $i \in S$ if and only if there is $n \in \mathbb{N}_0$ for which $P(X_n = i) > 0$. The set S can be either finite or countably infinite and we call it *state space* of the stochastic process $\{X_n, n \in \mathbb{N}_0\}$, its elements are *states of the process.* Without loss of generality, we will assume $S = \{0, 1 \dots, N\}$ or $S = \{0, 1, \dots\}$.

Definition 2.1. A sequence of integer-valued random variables $\{X_n, n \in \mathbb{N}_0\}$ is said to be *discrete time Markov chain* with the state space S, if

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \tag{2.1}$$

for all $n = 0, 1, \dots$ and all $i, j, i_{n-1}, \dots, i_0 \in S$ such that $P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$.

Relation (2.1) is so called *Markov property;* it means that the conditional probability distribution of a future state of the process (conditional on both past and present states) depends only on the present state, not on the sequence of events that preceded it.

Conditional probabilities

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

(if defined) are called *transition probabilities* from state i at time n to state j at time $n+1$, sometimes also *transition probabilities of the first order.*

Similarly, conditional probabilities

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

for $m \in \mathbb{N}$ (if defined) are called to be transition probabilities from state i at time n to state j at time $n+m$, or *transition probabilities of the order* $m-$. If the transition probabilities $p_{ij}(n, n+m)$ do not depend on times n and $n+m$, but on their difference, i.e., on m, only, we say that the Markov chain is *homogeneous.*

Consider a homogeneous Markov chain $\{X_n\}$. The first order transition probabilities $P(X_{n+1} = j | X_n = i)$, (if defined), are independent of n; we will denote them by p_{ij}. Since for every $i \in S$ there is $n \in \mathbb{N}_0$ such that $P(X_n = i) > 0$ and thus the conditional probability $P(X_{n+1} = j | X_n = i) = p_{ij}$ is defined for all $j \in S$, all these probabilities create a squared matrix $\boldsymbol{P} = \{p_{ij}, i, j \in S\}$. Obviously, for every $n \in \mathbb{N}_0$

$$p_{ij} \geq 0, \quad i, j \in S; \quad \sum_{j \in S} p_{ij} = 1, \quad i \in S. \tag{2.2}$$

Any squared matrix the elements of which satisfy condition (2.2) is called *stochastic matrix.* The matrix $\boldsymbol{P}$ of transition probabilities of a homogeneous Markov chain is the stochastic matrix.

Consider probabilities

$$p_i = P(X_0 = i), \quad i \in S.$$

Obviously,

$$p_i \geq 0, \quad i \in S; \quad \sum_{i \in S} p_i = 1. \tag{2.3}$$

The probability distribution $\boldsymbol{p} = \{p_i, i \in S\}$ is called to be the *initial distribution* of a Markov chain.

All the finite-dimensional distributions of a Markov chain are uniquely determined by the probabilities $P(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k)$ for all $k \in \mathbb{N}_0$ and all $i_k \in S$.

Theorem 2.1. *Let $\{X_n, n \in \mathbb{N}_0\}$ be a stochastic process with the state space $S = \{0, 1, \ldots\}$. Let $\boldsymbol{p} = \{p_i, i \in S\}$ be a vector that satisfies (2.3) and $\boldsymbol{P} = \{p_{ij}, i, j \in S\}$ be a matrix that satisfies (2.2). Then $\{X_n, n \in \mathbb{N}_0\}$ is a homogeneous Markov chain with the initial distribution $\boldsymbol{p}$ and the probability transition matrix $\boldsymbol{P}$ if and only if all the finite-dimensional distributions of this process are of the form*

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_k = i_k) = p_{i_0} p_{i_0 i_1} \ldots p_{i_{k-1} i_k} \tag{2.4}$$

for all $i_0, i_1, \ldots, i_k \in S$ and all $k \in \mathbb{N}_0$.

Proof. Let us recall properties of conditional probabilities: if $A_0, A_1, \dots, A_k$ are random events then

$$P\left(\bigcap_{i=0}^{k} A_i\right) = P\left(A_k \left| \bigcap_{i=0}^{k-1} A_i\right.\right).P\left(A_{k-1} \left| \bigcap_{i=0}^{k-2} A_i\right.\right) \dots P(A_1|A_0).P(A_0),$$

provided

$$P\left(\bigcap_{i=0}^{k-1} A_i\right) > 0.$$

Let $\{X_n\}$ be a homogeneous Markov chain. Put $A_j = [X_j = i_j], j = 0, 1, \dots, k$ and assume that

$$P(X_0 = i_0, \dots, X_{k-1} = i_{k-1}) > 0. \tag{2.5}$$

Then we get

$$\begin{aligned} &P(X_0 = i_0, \dots, X_k = i_k) \\ &\quad = P(X_k = i_k | X_{k-1} = i_{k-1}, \dots, X_0 = i_0) \dots P(X_1 = i_1 | X_0 = i_0) P(X_0 = i_0), \end{aligned}$$

and from here, due to the Markov property, we get (2.4). In case that (2.5) is not satisfied put

$$j^* = \min\{j \geq 0 : P(X_0 = i_0, \dots, X_j = i_j) = 0\}.$$

If $j^* = 0$, $P(X_0 = i_0) = p_{i_0} = 0$ and (2.4) holds. If $j^* > 0$, then

$$P(X_0 = i_0, \dots X_{j^*-1} = i_{j^*-1}) = p_{i_0} p_{i_0 i_1} \dots p_{i_{j^*-2} i_{j^*-1}} > 0$$

and

$$P(X_0 = i_0, \dots, X_{j^*} = i_{j^*}) = 0.$$

From here, since

$$p_{i_{j^*-1} i_{j^*}} = \frac{P(X_{j^*} = i_{j^*}, X_{j^*-1} = i_{j^*-1}, \dots, X_0 = i_0)}{P(X_{j^*-1} = i_{j^*-1}, \dots, X_0 = i_0)} = 0,$$

(2.4) follows.

Now, let us assume that the vector $\boldsymbol{p}$ satisfying (2.3) and the matrix $\boldsymbol{P}$ that satisfies (2.2) are given. Let us consider the process $\{X_n\}$ of integer-valued random variables the finite-dimensional distributions of that are given by (2.4). According to the Daniell-Kolmogorov theorem 1.1 such process exists. We will show that this process is a homogeneous Markov chain with the initial distribution $\boldsymbol{p}$ and the transition probability matrix $\boldsymbol{P}$.

Obviously, $P(X_0 = i_0) = p_{i_0}$ for every $i_0 \in S$.

Next, if $P(X_n = i) > 0$, then according to (2.4)

$$P(X_{n+1} = j | X_n = i) = \frac{P(X_{n+1} = j, X_n = i)}{P(X_n = i)}$$

$$= \frac{\sum_{i_0, i_1 \dots, i_{n-1}} P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i, X_{n+1} = j)}{\sum_{i_0, i_1 \dots, i_{n-1}} P(X_0 = i_0, \dots, X_n = i)}$$

$$= \frac{\sum_{i_0, i_1 \dots, i_{n-1}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i} p_{ij}}{\sum_{i_0, i_1 \dots, i_{n-1}} p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i}} = p_{ij}.$$

It remains to prove the Markov property of $\{X_n\}$.
Let

$$p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i} > 0.$$

Then

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \frac{p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i} p_{ij}}{p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i}} = p_{ij}.$$

□

Using properties of conditional probabilities and Theorem 2.1 we can also prove that

$$P(X_{m+n} = k_{m+n}, X_{m+n-1} = k_{m+n-1}, \dots, X_{m+1} = k_{m+1} | X_m = k_m, \dots, X_0 = k_0)$$
$$= P(X_{m+n} = k_{m+n}, X_{m+n-1} = k_{m+n-1}, \dots, X_{m+1} = k_{m+1} | X_m = k_m)$$

for all $k_i \in S, i = 0, 1, \dots, m+n,\ m, n \in \mathbb{N}$ and

$$P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_{n+1} = i_{n+1}, \dots, X_{n+m} = i_{n+m} | X_n = i_n)$$
$$= P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1} | X_n = i_n)$$
$$\times P(X_{n+1} = i_{n+1}, \dots, X_{n+m} = i_{n+m} | X_n = i_n)$$

for all $i_j \in S, j = 0, \dots, n+m,\ m, n \in \mathbb{N}$ a $P(X_n = i_n) > 0$. It means that given outcome X_n at present time n the future outcomes at times $n+1, \dots, n+m$ and past outcomes at times $n-1, \dots, 1, 0$ are mutually independent (conditional independence).

Remark. The assertion of Theorem 2.1 can be extended to a non homogeneous Markov chains. All the finite-dimensional distributions then will be determined by the initial distribution and the system of transition probabilities, i.e., we will get

$$P(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) = p_{i_0} p_{i_0 i_1}(0, 1) \dots p_{i_{k-1} i_k}(k-1, k)$$

for $i_0, i_1, \dots, i_k \in S, k \in \mathbb{N}_0$.

Now, let us again consider a homogeneous Markov chain with the transition probability matrix $\boldsymbol{P}$. Put $p_{ij}^{(0)} = \delta_{ij}$, where δ_{ij} denotes the Kronecker symbol

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

Next, put $p_{ij}^{(1)} = p_{ij}$ and for $n \in \mathbb{N}$ define

$$p_{ij}^{(n+1)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}. \tag{2.6}$$

We can prove that the series in (2.6) are convergent for every $n \geq 1$, since $p_{ij}^{(2)} \leq \sum_{k \in S} p_{ik} = 1$ and by using the mathematical induction, we get $p_{ij}^{(n)} \leq 1$. Similarly we can prove that the matrices $\boldsymbol{P}^{(n)}$ of elements $p_{ij}^{(n)}$ are stochastic matrices. From (2.6) we have

$$\boldsymbol{P}^{(2)} = \boldsymbol{P}.\boldsymbol{P} = \boldsymbol{P}^2 \text{ and generally } \boldsymbol{P}^{(n)} = \boldsymbol{P}^{(n-1)}.\boldsymbol{P} = \boldsymbol{P}.\boldsymbol{P}^{(n-1)} = \boldsymbol{P}^n.$$

Theorem 2.2. *Let $\{X_n\}$ be a homogeneous Markov chain with the transition probability matrix $\boldsymbol{P}$. Then the transition probabilities of order n satisfy*

$$P(X_{m+n} = j | X_m = i) = p_{ij}^{(n)}, \quad i, j \in S \tag{2.7}$$

for all integer $m \geq 0, n \geq 0$ and $P(X_m = i) > 0$.

Proof. For $n = 0, 1$ relation (2.7) is obvious. Let us assume that (2.7) holds for some integer $n > 1$ and all $i, j \in S$ and $m \geq 0$. Then, using the mathematical induction and the Markov property, we have

$$\begin{aligned} P(X_{m+n+1} = j | X_m = i) &= \sum_{k \in S} P(X_{m+n+1} = j, X_{m+n} = k | X_m = i) \\ &= \sum_{k \in S} P(X_{m+n} = k | X_m = i) P(X_{m+n+1} = j | X_{m+n} = k, X_m = i) \\ &= \sum_{k \in S} P(X_{m+n} = k | X_m = i) P(X_{m+n+1} = j | X_{m+n} = k) \\ &= \sum_{k \in S} p_{ik}^{(n)} p_{kj} = p_{ij}^{(n+1)}. \end{aligned}$$

(Here we have assumed that $P(X_{m+n} = k, X_m = i) > 0$ for all k. If $P(X_{m+n} = k, X_m = i) = 0$ for some k, then $P(X_{m+n+1} = j, X_{m+n} = k | X_m = i) = 0$ and also $P(X_{m+n} = k | X_m = i) = p_{ik}^{(n)} = 0$, thus $p_{ik}^{(n)} p_{kj} = 0$.) □

The relation in (2.6) can be easily generalized to the equation

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)} \tag{2.8}$$

for all $m, n \in \mathbb{N}_0$, that is called *the Chapmanov-Kolmogorov equation.* The matrix form of (2.8) is $\boldsymbol{P}^{(m+n)} = \boldsymbol{P}^{(m)} \boldsymbol{P}^{(n)}$.

The (unconditional) probabilities $p_j(n) = P(X_n = j)$ are said to be *absolute probabilities* at time n and satisfy

$$\begin{aligned} P(X_n = j) &= \sum_{k \in S} P(X_0 = k, X_n = j) = \sum_{k \in S} P(X_n = j | X_0 = k) P(X_0 = k) \\ &= \sum_{k \in S} p_k \; p_{kj}^{(n)}. \end{aligned}$$

If we denote $\boldsymbol{p}(n) = \{p_j(n), j \in S\}$, we can also write (all the vectors are columns)

$$\boldsymbol{p}(n)^T = \boldsymbol{p}^T \boldsymbol{P}^n, \quad n \in \mathbb{N}_0. \tag{2.9}$$

We can see that $\sum_{j \in S} p_j(n) = 1$ for all $n \in \mathbb{N}_0$, hence, $\{p_j(n), j \in S\}$ is the absolute distribution of the Markov chain at time n.

Remark. If the state space S is finite the elements of the matrix $\boldsymbol{P}^n$ can be determined by using the Perron formula (see Appendix B, Theorem B.6).

2.2 Examples of Markov chains

Example 2.1. A sequence $\{X_n, n \in \mathbb{N}_0\}$ of independent integer-valued random variables is the Markov chain since for every $n \in \mathbb{N}_0$ and integers $i, j, i_{n-1}, \dots, i_0$,

$$P(X_{n+1} = j | X_n = i, \dots, X_0 = i_0) = P(X_{n+1} = j) = P(X_{n+1} = j | X_n = i)$$

which is the Markov property (2.1). If X_n are moreover identically distributed with the distribution $\{a_i, i \in \mathbb{N}_0\}$, $\{X_n, n \in \mathbb{N}_0\}$ is the homogeneous Markov chain with the initial distribution

$$\boldsymbol{p} = (a_0, a_1, \dots)^T$$

and the transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} a_0 & a_1 & \dots \\ a_0 & a_1 & \dots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

Example 2.2. Let $\{Y_k, k \in \mathbb{N}\}$ be a sequence of independent integer-valued random variables. Define

$$X_0 = 0, \quad X_n = \sum_{k=1}^{n} Y_k, \quad n \geq 1.$$

We show that $\{X_n, n \in \mathbb{N}_0\}$ has the Markov property. From the definition of the conditional probability the left-hand side of (2.1) is

$$\frac{P(X_{n+1} = j, X_n = i, \ldots, X_1 = i_1, X_0 = 0)}{P(X_n = i, \ldots, X_1 = i_1, X_0 = 0)}. \tag{2.10}$$

Due to the equivalence of the events

$$[X_{n+1} = j, X_n = i, \ldots, X_1 = i_1, X_0 = 0], \quad [Y_{n+1} = j - i, Y_n = i - i_{n-1}, \ldots, Y_1 = i_1]$$

and

$$[X_n = i, \ldots, X_1 = i_1, X_0 = 0], \quad [Y_n = i - i_{n-1}, \ldots, Y_1 = i_1]$$

and due to the independence of random variables Y_n, the expression in (2.10) equals $P(Y_{n+1} = j - i)$. The right -hand side of (2.1) is

$$P(X_{n+1} = j | X_n = i) = \frac{P(Y_{n+1} = j - i, X_n = i)}{P(X_n = i)} = P(Y_{n+1} = j - i).$$

The sequence $\{X_n, n \in \mathbb{N}_0\}$ is the Markov chain with the transition probabilities $p_{ij}(n, n+1) = P(Y_{n+1} = j - i)$. In case that Y_n are identically distributed it is the homogeneous Markov chain. Especially, the random walk from Example 1.2 is the homogeneous Markov chain with the state space $S = \{0, \pm 1 \ldots\}$, and the transition probabilities $p_{i,i+1} = p_{i,i-1} = \frac{1}{2}, i \in S$.

Example 2.3. *Gambler's ruin problem* Two gamblers A and B play repeatedly a game that ends by either the winning of one of them. In each round, A wins 1 dollar with probability $0 < p < 1$ or loses 1 dollar with probability $q = 1 - p$ (and equivalently, B wins 1 dollar with probability q and loses 1 dollar with probability p). The initial fortune of A is z dollars while B has $a - z$ dollars, where a is the total amount of money at the game. The rounds are repeated independently on the past rounds and the game stops when either player has no money (he is ruined while the other one wins the total).

Let X_n be the fortune of the player A after the $n-$th round. Then we can show that $\{X_n\}$ is a homogeneous Markov chain with the state space $S = \{0, 1, \ldots, a\}$, the initial distribution $p_z = 1, p_j = 0, j \neq z$ and the transition probabilities $p_{00} = p_{aa} = 1,\ p_{i,i+1} = p,\ p_{i,i-1} = q, 1 \leq i \leq a-1$. The transition probability matrix thus is

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ q & 0 & p & 0 & \dots & 0 & 0 & 0 \\ 0 & q & 0 & p & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & q & 0 & p \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

Other interpretation of the problem can be *a random walk with absorbing barriers*. We consider a particle moving on the integer-valued grid between barriers at $x = 0$ and $x = a > 0$, at any time one unit step to the right with probability $p > 0$ or to the left with probability $q = 1-p$, independently on the previous steps. The barriers 0 and a are absorbing; when either one of them is reached the particle remains there. The sequence of random variables describing the movement of the particle is again the Markov chain as above.

Example 2.4. The Galton-Watson branching process from Example 1.3) is a Markov chain. Number of individuals in the null-th generation is $X_0 = 1$, number of individuals in the first generation is a random variable X_1 with the distribution $P(X_1 = j) = a_j, j = 0, 1, \dots$ and the number of individuals in the $n-$th generation is $X_n = U_{n1} + \dots + U_{nX_{n-1}}$, where U_{ni} are random variables with the same distribution as X_1, independent of each other and of $X_{n-1}, X_{n-2}, \dots, X_0$. Then

$$\begin{aligned} &P(X_n = j | X_{n-1} = i, \dots, X_1 = i_1, X_0 = 1) \\ &\quad = P\left(\sum_{k=1}^{X_{n-1}} U_{nk} = j \,\middle|\, X_{n-1} = i, \dots, X_1 = i_1, X_0 = 1 \right) \\ &\quad = P\left(\sum_{k=1}^{i} U_{nk} = j \,\middle|\, X_{n-1} = i, \dots, X_1 = i_1, X_0 = 1 \right) \\ &\quad = P\left(\sum_{k=1}^{i} U_{nk} = j \right) = P(X_n = j | X_{n-1} = i), \end{aligned}$$

which is the Markov property.

The initial distribution of the process $\{X_n, n \in \mathbb{N}_0\}$ is $\boldsymbol{p} = (0, 1, 0 \dots)^T$, the transition probabilities are

$$P(X_n = j | X_{n-1} = i) = P\left(\sum_{k=1}^{i} U_{nk} = j \right) = a_j^{*i},$$

where $\{a_j\}^{*i}$ is the $i-$th convolution of $\{a_j\}$ (see Appendix A.)

Example 2.5. *Generation of a Markov chain* with the given state space S, the initial distribution $\boldsymbol{p} = \{p_i, i \in S\}$ and the transition probability matrix $\boldsymbol{P} = \{p_{ij}, i, j \in S\}$.

Let $\{U_n, n \in \mathbb{N}_0\}$ be a sequence of independent identically distributed random variables with the uniform distribution on the interval $[0, 1]$. Define the random variable X_0 by

$$X_0 = k \Longleftrightarrow \sum_{i=0}^{k-1} p_i < U_0 \leq \sum_{i=0}^{k} p_i$$

(put $\sum_{i=0}^{-1} p_i = 0$). Then X_0 takes values k with the probabilities p_k. Further, define the function $f(i, u)$ on $S \times [0, 1]$ by

$$f(i, u) = k \Longleftrightarrow \sum_{j=0}^{k-1} p_{ij} < u \leq \sum_{j=0}^{k} p_{ij}$$

and random variables X_n by a recursion

$$X_{n+1} = f(X_n, U_{n+1}), \quad n \geq 0.$$

Given $X_n = i$, the random variable X_{n+1} takes value k with the probability p_{ik}. We can see that X_n is a function of $U_n, U_{n-1}, \ldots, U_0$, only, thus X_n and U_{n+1} are independent and

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

Moreover,

$$\begin{aligned} &P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \\ &= P(f(i, U_{n+1}) = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(f(i, U_{n+1}) = j) = p_{ij}, \end{aligned}$$

since $X_n, X_{n-1}, \ldots, X_0$ are independent of U_{n+1}. The sequence $\{X_n, n \in \mathbb{N}_0\}$ has the Markov property.

Example 2.6. *A storage model.* A store of a good (in integer-valued units) is checked at given times $t_0, t_1, t_2 \ldots$ Demand for the good at interval $[t_n, t_{n+1})$ is a random variable D_n. Suppose that $D_n, n = 0, 1, \ldots$ are independent identically distributed integer-valued random variables such that $P(D_n = k) = p_k, k = 0, 1, \ldots.$ The replenishment strategy is as follows: Let m, M be integers such that $m < M$. If the store X_n at time t_n is such that $m < X_n \leq M$, there is no replenishment, if $X_n \leq m$, the stock is replenished to the level M. Suppose that D_n is independent of X_0 and $X_0 \leq M$. Then the stock at time t_{n+1} (before possible replenishment) is

$$X_{n+1} = \begin{cases} \max\,(X_n - D_n, 0), & m < X_n \leq M, \\ \max\,(M - D_n, 0), & X_n \leq m. \end{cases}$$

we see that X_{n+1} depends on X_n and D_n, only, and X_n, D_n are independent. It means that $\{X_n, n \in \mathbb{N}_0\}$ is the Markov chain with states $0, 1, \dots, M$. The transition probability matrix has elements

$$\begin{aligned}
&p_{i0} = q_M, && i = 0,1,\dots,m, && \\
&p_{ij} = p_{M-j}, && i = 0,1,\dots,m, && j = 1,\dots,M, \\
&p_{i0} = q_i, && i = m+1,\dots,M, && \\
&p_{ij} = p_{i-j}, && i = m+1,\dots,M, && j = 1,\dots,i, \\
&p_{ij} = 0, && i = m+1,\dots,M, && j = i+1,\dots,M,
\end{aligned}$$

where

$$q_k = p_k + p_{k+1} + \dots, \qquad m < k \le M.$$

Example 2.7. *A car insurance model.* An insurance company offers three-level discount of a car premium: 0%, 30% and 50%. Each year the level of the premium of a policy-holder is determined on the basis of the level in the previous year and the number of reported claims in that year. In the case of a claim-free year, a policy-holder moves to the immediately higher discount level. A claim causes the policy-holder to move one step-back level if exactly one claim has been reported, and two step-back level if more then one claim has appeared. Assume that the number of claims in the $n-$th year is a random variable Y_n, and further assume that $Y_n, n = 1, 2, \dots$ are independent and identically distributed and follow the Poisson distribution with parameter λ. Let X_n denote the discount level in the $n-$th year and assume that X_1 and Y_1 are independent.

Obviously, for $n \ge 1$

$$X_{n+1} = \begin{cases} \min\,(X_n + 1, 2) & \text{pro } Y_n = 0, \\ \max\,(X_n - 1, 0) & \text{pro } Y_n = 1, \\ 0 & \text{pro } Y_n > 1. \end{cases}$$

We can conclude that $\{X_n, n \in \mathbb{N}\}$ is the Markov chain with the state space $S = \{0, 1, 2\}$, with the initial distribution $\boldsymbol{p} = (1, 0, 0)^T$ and the transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} 1 - e^{-\lambda} & e^{-\lambda} & 0 \\ 1 - e^{-\lambda} & 0 & e^{-\lambda} \\ 1 - e^{-\lambda} - \lambda e^{-\lambda} & \lambda e^{-\lambda} & e^{-\lambda} \end{pmatrix}.$$

2.3 Classification of states

We start with the definition of a stopping time.

Definition 2.2. Let $\{X_n, n \in \mathbb{N}_0\}$ be a stochastic process with a countable state space defined on $(\Omega, \mathcal{A}, P)$. A random variable $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ is called to be a *stopping time* of $\{X_n, n \in \mathbb{N}_0\}$ if the events $[\tau = n]$ ($[\tau \le n]$, respectively,) belong to the σ-field $\mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\}$ generated by random variables $X_0, X_1, \dots, X_n$.

Theorem 2.3. *Let $\{X_n, n \in \mathbb{N}_0\}$ be a homogeneous Markov chain on $(\Omega, \mathcal{A}, P)$ with a state space S and a transition probability matrix $\boldsymbol{P}$. Let τ be a stopping time of the process $\{X_n, n \in \mathbb{N}_0\}$ such that $P(\tau < \infty) = 1$. Then*

$$P(X_{\tau+1} = j | X_\tau = i, X_{\tau-1} = i_{\tau-1}, \dots, X_0 = i_0) = P(X_{\tau+1} = j | X_\tau = i) = p_{ij}, \; i, j \in S, \tag{2.11}$$

where $X_\tau(\omega) = X_{\tau(\omega)}(\omega)$ for all $\omega \in \Omega$.

Proof. Let $K_\tau = \{k \in \mathbb{N}_0 : [X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0] \subset [\tau = k]\}$. Due to the assumption $P(\tau < \infty) = 1$ and the Markov property we have

$$\begin{aligned}
&P(X_{\tau+1} = j, X_\tau = i, X_{\tau-1} = i_{\tau-1}, \dots, X_0 = i_0) \\
&\quad = \sum_{k=0}^{\infty} P(X_{\tau+1} = j, X_\tau = i, X_{\tau-1} = i_{\tau-1}, \dots, X_0 = i_0, \tau = k) \\
&\quad = \sum_{k=0}^{\infty} P(X_{k+1} = j, X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0, \tau = k) \\
&\quad = \sum_{k \in K_\tau} P(X_{k+1} = j, X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0, \tau = k) \\
&\quad = \sum_{k \in K_\tau} P(X_{k+1} = j, X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0) \\
&\quad = \sum_{k \in K_\tau} [P(X_{k+1} = j | X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0) \\
&\qquad\qquad \times P(X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0)] \\
&\quad = \sum_{k \in K_\tau} P(X_{k+1} = j | X_k = i) P(X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0) \\
&\quad = p_{ij} \sum_{k \in K_\tau} P(X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0) \\
&\quad = p_{ij} \sum_{k=0}^{\infty} P(X_k = i, X_{k-1} = i_{k-1}, \dots, X_0 = i_0, \tau = k) \\
&\quad = p_{ij} P(X_\tau = i, X_{\tau-1} = i_{\tau-1}, \dots, X_0 = i_0).
\end{aligned}$$

Similarly we get

$$P(X_{\tau+1} = j, X_\tau = i) = \sum_{k=0}^{\infty} P(X_{\tau+1} = j, X_\tau = i, \tau = k) = p_{ij} P(X_\tau = i).$$

□

Theorem 2.3 says us that the Markov property holds true when the process $\{X_n, n \in \mathbb{N}_0\}$ is observed in random times $\tau + m, m \in \mathbb{N}$, where τ is the stopping time. This property is called *strong Markov property.* If $\tau < \infty$ and $X_\tau = i$ for $i \in S$, then it can be shown that $(X_0, \dots, X_{\tau-1})$ a $(X_{\tau+1}, \dots, X_{\tau+m}), m = 1, 2, \dots$ are conditionally independent.

In the sequel, we will consider homogeneous Markov chains only. Further we will use the following notation:

$$P(.|X_0 = j) = P_j(.)$$

and given $X_0 = j$, the conditional mean and conditional variance will be accordingly denoted by E_j and Var_j.

Put $\tau_j(0) = 0$ and define

$$\tau_j(1) = \inf\{n > 0 : X_n = j\} \tag{2.12}$$

where $\inf\{\emptyset\} = \infty$. According to this definition, $\tau_j(1)$ is a random variable taking values $1, 2, \dots,$ or ∞ and denotes the random time in which the chain visits state j for the first time. It is called *the first passage time to state j*. Similarly, we can define times of next visits to state j :

$$\tau_j(k+1) = \inf\{n > \tau_j(k) : X_n = j\},\ k = 1, 2, \dots, \tag{2.13}$$

when $\tau_j(k) < \infty$.

The first passage time $\tau_j(1)$ is the stopping time that satisfies definition 2.2 since

$$[\tau_j(1) = n] = [X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j] \in \mathcal{F}_n = \sigma\{X_0, X_1, \dots, X_n\}$$

for $n = 1, 2, \dots,\ j \in S$. Similarly, random variables $\tau_j(k)$ for $k = 2, 3, \dots$ are stopping times.

Definition 2.3. State j of a Markov chain is said to be *recurrent* if, starting at state j, the chain will return to j with probability 1, i.e., in a finite number of transitions, in other words,

$$P_j(\tau_j(1) < \infty) = 1.$$

State j is said to be *transient,* if

$$P_j(\tau_j(1) = \infty) > 0$$

which means that there is a positive probability that the chain will leave state j forever.

Definition 2.4. The recurrent state j is called *positive recurrent* if $\mu_j := E_j\tau_j(1) < \infty$, and *null recurrent* if $\mu_j = \infty$.

Now, let us introduce the following notation:

$$\begin{aligned} f_{ij}^{(0)} &= 0, \\ f_{ij}^{(n)} &= P_i(\tau_j(1) = n), \quad n \geq 1, \\ f_{ij} &= \sum_{n=1}^{\infty} f_{ij}^{(n)} = P_i(\tau_j(1) < \infty). \end{aligned}$$

We can see that j is recurrent if $f_{jj} = 1$, positive recurrent if $\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)}$ is convergent, and null recurrent if this series is divergent. If $f_{jj} < 1$, state j is transient.

Theorem 2.4. *Let $p_{ij}^{(n)}$ be transition probabilities of order n. Then*

$$p_{jj}^{(n)} = \sum_{k=0}^{n} f_{jj}^{(k)} p_{jj}^{(n-k)}, \quad n \geq 1, \tag{2.14}$$

$$p_{ij}^{(n)} = \sum_{k=0}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}, \quad n \geq 0, \quad i \neq j. \tag{2.15}$$

The generating functions P_{ij}, F_{ij} of sequences $\{p_{ij}^{(n)}\}, \{f_{ij}^{(n)}\}$ satisfy

$$P_{jj}(s) = \frac{1}{1 - F_{jj}(s)}, \quad |s| < 1, \tag{2.16}$$

$$P_{ij}(s) = F_{ij}(s) P_{jj}(s), \quad |s| < 1 \quad i \neq j. \tag{2.17}$$

Proof. Since $[X_n = j] \subset [\tau_j(1) \leq n]$, we have

$$\begin{aligned} p_{jj}^{(n)} &= P(X_n = j | X_0 = j) = P_j(X_n = j) = \sum_{k=1}^{n} P_j(X_n = j | \tau_j(1) = k) P_j(\tau_j(1) = k) \\ &= \sum_{k=1}^{n} P_j(X_n = j | \tau_j(1) = k) f_{jj}^{(k)} \end{aligned}$$

and by using the Markov property we get

$$\begin{aligned} P_j(X_n = j | \tau_j(1) = k) &= P(X_n = j | X_0 = j, X_1 \neq j, \ldots, X_{k-1} \neq j, X_k = j) \\ &= P(X_n = j | X_k = j) = p_{jj}^{(n-k)}. \end{aligned}$$

From here (2.14) follows, since $f_{jj}^{(0)} = 0$. The proof of (2.15) is analogous.

Since $P_{jj}(s) = \sum_{n=0}^{\infty} p_{jj}^{(n)} s^n,\ F_{jj}(s) = \sum_{n=0}^{\infty} f_{jj}^{(n)} s^n$ are generating functions, we obtain

$$P_{jj}(s) = \sum_{n=0}^{\infty} p_{jj}^{(n)} s^n = 1 + \sum_{n=1}^{\infty} p_{jj}^{(n)} s^n = 1 + \sum_{n=1}^{\infty} s^n \left(\sum_{k=0}^{n} f_{jj}^{(k)} p_{jj}^{(n-k)} \right) = 1 + F_{jj}(s) P_{jj}(s),$$

which results in (2.16). Formula (2.17) can be obtained in the same way.

□

Theorem 2.5. *State j is recurrent if and only if*

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty$$

and transient if and only if

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty.$$

Proof. Obviously $f_{jj} = \lim_{s \to 1-} F_{jj}(s) = F_{jj}(1)$. State j is recurrent if and only if $f_{jj} = 1$, equivalently, according to (2.16) if and only if

$$\sum_{n=0}^{\infty} p_{jj}^{(n)} = \lim_{s \to 1-} P_{jj}(s) = \lim_{s \to 1-} \frac{1}{1 - F_{jj}(s)} = \infty.$$

□

Remark. We have used the convention $\lim_{s \to 1-} A(s) = A(1)$ where A is a generating function. We will use the same convention for derivatives of a generating function.

Example 2.8. Consider a sequence of independent tosses of a dice. Let X_m denotes the maximum number achieved up to $m-$th toss. Then $\{X_m\}$ is the homogeneous Markov chain with the state space $S = \{1, \dots, 6\}$ and transition probabilities

$$p_{ij} = \begin{cases} \frac{1}{6}, & i < j, \\ \frac{j}{6}, & i = j, \\ 0, & i > j. \end{cases}$$

Transition probabilities of order n are

$$p_{ij}^{(n)} = \begin{cases} \left(\frac{j}{6}\right)^n - \left(\frac{j-1}{6}\right)^n, & i < j, \\ \left(\frac{j}{6}\right)^n, & i = j, \\ 0, & i > j. \end{cases}$$

We can see that $\sum_{n=0}^{\infty} p_{jj}^{(n)}$ is convergent for $j = 1, \ldots, 5$ and divergent for $j = 6$. States $1, \ldots, 5$ are transient, state 6 is recurrent.

Example 2.9. *A symmetric random walk on* $\mathbb{Z}$ is the Markov chain with the state space $S = \{0, \pm 1, \ldots\}$ and transition probabilities $p_{i,i+1} = p_{i,i-1} = \frac{1}{2}, i \in S$ (see Examples 1.2 a 2.2). Transition probabilities from state j to state j after n steps are

$$p_{jj}^{(n)} = \begin{cases} \binom{2k}{k} \left(\frac{1}{2}\right)^{2k}, & n = 2k, \quad k = 0, 1, \ldots, \\ 0, & n = 2k+1, \quad k = 0, 1, \ldots. \end{cases}$$

When using the Stirling formula

$$k! \sim k^k \mathrm{e}^{-k} \sqrt{2\pi k}, \quad k \to \infty,$$

we get for all $j \in S$ and $k > 0$

$$p_{jj}^{(2k)} \sim \frac{1}{\sqrt{\pi k}}, \quad k \to \infty.$$

From here it follows that $\sum_{n=0}^{\infty} p_{jj}^{(n)} = \infty$ for all $j \in S$, therefore, all states are recurrent.

Now, let us consider visiting times $\tau_j(k)$ defined in (2.13). The random variables $T_1 = \tau_j(1), T_2 = \tau_j(2) - \tau_j(1), T_3 = \tau_j(3) - \tau_j(2), \ldots$ are times between returns to state j. They satisfy the following theorem.

Theorem 2.6. *Under condition that $\tau_j(1) < \infty, \ldots, \tau_j(k) < \infty$, $T_1, T_2, \ldots, T_k$ are independent random variables. If the chain starts from state j, $T_1, T_2, \ldots, T_k$ have the same distribution given by probabilities $\{f_{jj}^{(n)}\}$. If the chain starts from state i, T_1 has distribution $\{f_{ij}^{(n)}\}$ and $T_2, \ldots, T_k$ distribution $\{f_{jj}^{(n)}\}$.*

Proof. For simplicity, we prove the theorem only for random variables T_1, T_2. Notice that

$$
\begin{aligned}
[T_1 = m, T_2 = n] &= [\tau_j(1) = m, \tau_j(2) = m+n] \\
&= [X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j, X_{m+1} \neq j, \dots, X_{m+n-1} \neq j, X_{m+n} = j].
\end{aligned}
$$

Then

$$
\begin{aligned}
&P_j(T_1 = m, T_2 = n) = P(T_1 = m, T_2 = n | X_0 = j) \\
&= P(X_{m+n} = j, X_{m+n-1} \neq j, \dots, X_{m+1} \neq j | X_m = j, X_{m-1} \neq j, \dots, X_1 \neq j, X_0 = j) \\
&\times P_j(X_m = j, X_{m-1} \neq j, \dots, X_1 \neq j).
\end{aligned}
$$

Now we utilize a simple generalization of the Markov property, homogeneity of the chain and relation 2.4 and get

$$
\begin{aligned}
P&\left(X_{m+n} = j, X_{m+n-1} \neq j, \dots, X_{m+1} \neq j \middle| X_m = j, X_{m-1} \neq j, \dots, X_1 \neq j, X_0 = j\right) \\
&= P(X_{m+n} = j, X_{m+n-1} \neq j, \dots, X_{m+1} \neq j | X_m = j) \\
&= P(X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j | X_0 = j) = P_j(\tau_j(1) = n).
\end{aligned}
$$

We can summarize that

$$
P_j(T_1 = m, T_2 = n) = P_j(\tau_j(1) = m) P_j(\tau_j(1) = n) = P_j(T_1 = m) P_j(T_1 = n). \quad (2.18)
$$

From here

$$
P_j(T_2 = n) = \sum_{m=1}^{\infty} P_j(T_1 = m, T_2 = n) = P_j(T_1 = n),
$$

thus, T_1, T_2 have the same distribution given by

$$
P_j(T_1 = k) = f_{jj}^{(k)}, \quad k \geq 1. \quad (2.19)
$$

It also follows from (2.18) that T_1, T_2 are independent.

Provided that the chain starts from state $i \neq j$, using similar arguments we get

$$
P_i(T_1 = m, T_2 = n) = P_i(\tau_j(1) = m) P_j(\tau_j(1) = n) = P_i(T_1 = m) P_j(T_1 = n),
$$

$$
P_i(T_2 = n) = \sum_{m=1}^{\infty} P_i(T_1 = m, T_2 = n) = P_j(T_1 = n).
$$

We can see that T_1, T_2 are independent, random variable T_1 has distribution

$$
P_i(T_1 = k) = P_i(\tau_j(1) = k) = f_{ij}^{(k)}, \quad k \geq 1, \quad (2.20)
$$

while T_2 has distribution

$$
P_i(T_2 = k) = P_j(T_1 = k) = f_{jj}^{(k)}, \quad k \geq 1. \quad (2.21)
$$

□

Now let us consider random variable N_j, that denotes the number of visits to state j after leaving the initial state, i.e.,

$$N_j = \sum_{n=1}^{\infty} I(X_n = j).$$

Obviously, N_j is a random variable taking values $k = 0, 1, \dots$. The distribution of this random variable is given in the following theorem.

Theorem 2.7. *It holds*

$$P_i(N_j = k) = \begin{cases} 1 - f_{ij}, & k = 0, \\ f_{ij} f_{jj}^{k-1}(1 - f_{jj}), & k > 0. \end{cases}$$

Proof. For $k = 0$ we have

$$P_i(N_j = 0) = 1 - P_i(N_j \geq 1) = 1 - P_i(\tau_j(1) < \infty) = 1 - f_{ij}.$$

For $k > 0$ we obtain

$$P_i(N_j \geq k) = P_i(\tau_j(k) < \infty) = P_i(T_1 + T_2 + \cdots + T_k < \infty).$$

It follows from Theorem 2.6 that the random variables $T_1, \dots, T_k$ are independent, T_1 with the distribution $\{f_{ij}^{(n)}\}$ while $T_2, \dots, T_k$ with the distribution $\{f_{jj}^{(n)}\}$. The generating function of the random variable $\tau_j(k)$ is $F_{ij}(s)[F_{jj}(s)]^{k-1}$. From here

$$P_i(N_j \geq k) = P_i(\tau_j(k) < \infty) = \sum_{r=0}^{\infty} P_i(\tau_j(k) = r) = F_{ij}(1)[F_{jj}(1)]^{k-1} = f_{ij} {f_{jj}}^{k-1}.$$

Thus,

$$P_i(N_j = k) = P_i(N_j \geq k) - P_i(N_j \geq k+1) = f_{ij} f_{jj}^{k-1} - f_{ij} f_{jj}^{k} = f_{ij} f_{jj}^{k-1}(1 - f_{jj}).$$

□

Theorem 2.8. *If j is a recurrent state of a Markov chain, then*

$$P_j(N_j = \infty) = 1.$$

If j is a transient state, then

$$P_i(N_j = \infty) = 0, \quad i \in S$$

and

$$P_j(N_j = k) = (1 - f_{jj})f_{jj}^k, \quad k \geq 0 \quad \textit{(geometric distribution)} .$$

Proof. Due to monotonicity of the random events

$$[N_j \geq k] \supset [N_j \geq k+1] \supset \dots$$

it holds

$$P_i(N_j = \infty) = \lim_{k\to\infty} P_i(N_j \geq k) = \lim_{k\to\infty} f_{ij} f_{jj}^{k-1}.$$

Suppose that j is recurrent. Then $f_{jj} = 1$, hence $P_j(N_j = \infty) = 1$. If j is transient, then $f_{jj} < 1$ and therefore $P_i(N_j = \infty) = 0$. The rest of the proof follows from Theorem 2.7. □

The previous results show us that if the state j is recurrent, then the chain visits it infinitely many times with probability one, while the number of visits to a transient state j is finite with probability one.

Now, let us introduce the following classification.

Definition 2.5. Let d_j be the greatest common divisor of numbers $n \geq 1$, such that $p_{jj}^{(n)} > 0$. If $d_j > 1$, we call state j to be *periodic with the period* d_j, if $d_j = 1$, we say that the state j *aperiodic.* If $p_{jj}^{(n)} = 0$ for all $n > 0$ we put $d(j) = 1$.

Example 2.10. In the gambler's ruin problem (Example 2.3) states $0,\ a$ are aperiodic and states $1, \dots, a-1$ periodic with period 2. In Example 2.9 (symmetric random walk on $\mathbb{Z}$) all states are periodic with period 2.

Remark. If $p_{jj} > 0$, holds, the state j is aperiodic. However, this condition is not necessary, for example, in the chain with the transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}$$

we have $p_{11} = 0,\ p_{11}^{(2)} = \frac{1}{2},\ p_{11}^{(3)} = \frac{1}{2}$, thus, $d_1 = 1$.

Definition 2.6. A positive recurrent and aperiodic state j is called *ergodic*.

Next, we will treat an asymptotic behaviour of transition probabilities but first we introduce the following lemma.

Lemma 2.1. Let $\{f_n, n \in \mathbb{N}_0\}$ be a sequence of real numbers such that $f_0 = 0,\ f_n \geq 0,\ \sum_{n=0}^{\infty} f_n = 1$ and the greatest common divisor of numbers n, such that $f_n > 0$ equals 1. Let us define the sequence $\{u_n\}$ by

$$\begin{aligned} u_0 &= 1, \\ u_n &= \sum_{k=0}^{n} f_k u_{n-k}, \quad n \geq 1. \end{aligned}$$

Let $\mu = \sum_{k=0}^{\infty} k f_k$. Then as $n \to \infty$

$$\begin{aligned} u_n &\to \frac{1}{\mu}, \quad \mu < \infty, \\ u_n &\to 0, \quad \mu = \infty. \end{aligned}$$

.

Proof. Feller (1964), pp. 331–333. □

Theorem 2.9. **(i)** *Let j be a transient state of a Markov chain. Then as $n \to \infty$*

$$p_{ij}^{(n)} \to 0, \quad i \in S.$$

(ii) *Let j be positive recurrent and aperiodic. Then as $n \to \infty$*

$$\begin{aligned} p_{jj}^{(n)} &\to \frac{1}{\mu_j}, \\ p_{ij}^{(n)} &\to \frac{f_{ij}}{\mu_j}, \quad i \neq j, \end{aligned}$$

where $\mu_j = \sum_{n=0}^{\infty} n f_{jj}^{(n)}$ is the mean value of the first visit to j.

(iii) *Let j be positive recurrent and periodic with period d_j. Then as $k \to \infty$*

$$\begin{aligned} p_{jj}^{(kd_j)} &\to \frac{d_j}{\mu_j}, \\ p_{ij}^{(kd_j+l)} &\to \frac{d_j}{\mu_j} \sum_{\nu=0}^{\infty} f_{ij}^{(d_j\nu+l)}, \quad 0 \leq l < d_j, \quad i \neq j \end{aligned}$$

and further, as $n \to \infty$

$$\overline{p}_{ij}^{(n)} := \frac{1}{n} \sum_{\nu=1}^{n} p_{ij}^{(\nu)} \to \frac{f_{ij}}{\mu_j}, \quad i \neq j.$$

(iv) *Let j be null recurrent. Then as $n \to \infty$*

$$p_{ij}^{(n)} \to 0, \quad i \in S.$$

Proof. (i) Let j be transient. Then $\sum_{n=0}^{\infty} p_{jj}^{(n)} < \infty$ (Theorem 2.5) and thus, $p_{jj}^{(n)} \to 0$ as $n \to \infty$. Similarly, according to (2.17) and (2.16)

$$\sum_{n=0}^{\infty} p_{ij}^{(n)} = P_{ij}(1) = \frac{F_{ij}(1)}{1 - F_{jj}(1)} = \frac{f_{ij}}{1 - f_{jj}} < \infty,$$

so that $p_{ij}^{(n)} \to 0, \quad i \neq j$.

(ii) Let j be positive recurrent and aperiodic. Put $f_n = f_{jj}^{(n)}, u_n = p_{jj}^{(n)}$ and notice that, if number 1 is the greatest common divisor of $\{n \geq 1 : p_{jj}^{(n)} > 0\}$, it is also the greatest common divisor of $\{n \geq 1 : f_{jj}^{(n)} > 0\}$. Thus, according to Lemma 2.1, $p_{jj}^{(n)} \to \frac{1}{\mu_j}$, where $\mu_j = \sum_{n=1}^{\infty} n f_{jj}^{(n)} = F'_{jj}(1)$.

For $i \neq j$ we get from (2.15) and for $n > N$

$$p_{ij}^{(n)} = \sum_{k=0}^{N} f_{ij}^{(k)} p_{jj}^{(n-k)} + \sum_{k=N+1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}, \tag{2.22}$$

hence we can write

$$\left| p_{ij}^{(n)} - \frac{f_{ij}}{\mu_j} \right| \leq \left| p_{ij}^{(n)} - \sum_{k=0}^{N} f_{ij}^{(k)} p_{jj}^{(n-k)} \right| + \left| \sum_{k=0}^{N} f_{ij}^{(k)} p_{jj}^{(n-k)} - \frac{1}{\mu_j} \sum_{k=0}^{N} f_{ij}^{(k)} \right| \tag{2.23}$$
$$+ \left| \frac{1}{\mu_j} \sum_{k=0}^{N} f_{ij}^{(k)} - \frac{f_{ij}}{\mu_j} \right|.$$

Podle (2.22) je

$$\left| p_{ij}^{(n)} - \sum_{k=0}^{N} f_{ij}^{(k)} p_{jj}^{(n-k)} \right| = \sum_{k=N+1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \leq \sum_{k=N+1}^{n} f_{ij}^{(k)} \leq \sum_{k=N+1}^{\infty} f_{ij}^{(k)} < \varepsilon$$

for $\varepsilon > 0$ and sufficiently large N ($N > N_0(\varepsilon)$), since $\sum_{k=1}^{\infty} f_{ij}^{(k)} = f_{ij} \leq 1 < \infty$.

The second term on the right-hand side of (2.23) for given N and $n > N + n_0(\varepsilon)$ is

$$\left| \sum_{k=0}^{N} f_{ij}^{(k)} p_{jj}^{(n-k)} - \frac{1}{\mu_j} \sum_{k=0}^{N} f_{ij}^{(k)} \right| \leq \sum_{k=0}^{N} f_{ij}^{(k)} \left| p_{jj}^{(n-k)} - \frac{1}{\mu_j} \right| < \varepsilon \sum_{k=0}^{N} f_{ij}^{(k)} \leq \varepsilon \sum_{k=0}^{\infty} f_{ij}^{(k)} \leq \varepsilon,$$

since we have proved that $p_{jj}^{(n)} \to \frac{1}{\mu_j}$ as $n \to \infty$.

Finally we have

$$\frac{1}{\mu_j} \left| \sum_{k=0}^{N} f_{ij}^{(k)} - f_{ij} \right| = \frac{1}{\mu_j} \sum_{k=N+1}^{\infty} f_{ij}^{(k)} < \frac{\varepsilon}{\mu_j}$$

for $N > N_0(\varepsilon)$. The expression on the left-hand side of (2.23) can be done arbitrarily small by a proper choice of n and N.

(iii) Let j be positive recurrent with period d_j. Then $p_{jj}^{(n)} = 0$, $f_{jj}^{(n)} = 0$ for $n \neq kd_j$, $k = 0, 1, \ldots$. Denote

$$\widetilde{p}_{jj}^{(k)} = p_{jj}^{(kd_j)}, \quad \widetilde{f}_{jj}^{(k)} = f_{jj}^{(kd_j)}.$$

When we apply (ii) to the sequence $\{\widetilde{p}_{jj}^{(k)}\}$ we get, as $k \to \infty$

$$\widetilde{p}_{jj}^{(k)} \to \frac{1}{\widetilde{\mu}_j},$$

where $\widetilde{\mu}_j = \sum_{k=1}^{\infty} k \widetilde{f}_{jj}^{(k)} = \widetilde{F}'_{jj}(1)$ and $\widetilde{F}_{jj}$ is the generating function of the sequence $\{\widetilde{f}_{jj}^{(k)}\}$. We have

$$\widetilde{F}_{jj}(s) = \sum_{k=0}^{\infty} \widetilde{f}_{jj}^{(k)} s^k = \sum_{k=0}^{\infty} f_{jj}^{(kd_j)} \left(s^{1/d_j}\right)^{kd_j} = \sum_{n=0}^{\infty} f_{jj}^{n} \left(s^{1/d_j}\right)^{n} = F_{jj}(s^{1/d_j}),$$

and from here

$$\widetilde{\mu}_j = \widetilde{F}'_{jj}(1) = \frac{1}{d_j} F'_{jj}(1) = \frac{\mu_j}{d_j},$$

thus

$$p_{jj}^{(kd_j)} \to \frac{d_j}{\mu_j}.$$

For $i \neq j$ we have from(2.15)

$$p_{ij}^{(kd_j+l)} = \sum_{\nu=0}^{k} f_{ij}^{(\nu d_j + l)} p_{jj}^{(k-\nu)d_j},$$

and further we can proceed analogously as in (ii).

According to the Tauber theorem, see (iv) in Appendix A, and from (2.17), and (2.16)

$$\lim_{n\to\infty} \overline{p}_{ij}^{(n)} = \lim_{n\to\infty} \frac{1}{n} \sum_{\nu=1}^{n} p_{ij}^{(\nu)} = \lim_{s\to 1-} (1-s) P_{ij}(s)$$

$$= \lim_{s\to 1-} \frac{1-s}{1-F_{jj}(s)} F_{ij}(s) = \frac{F_{ij}(1)}{F'_{jj}(1)} = \frac{f_{ij}}{\mu_j}.$$

(iv) If j is null recurrent and aperiodic, then by Lemma 2.1 $p_{jj}^{(n)} \to 0$ and from here $p_{ij}^{(n)} \to 0$ for $i \neq j$ analogously as in (ii), since every summands in (2.22) can be arbitrarily small for sufficiently large n and N. If j is null recurrent and periodic with the period d_j, then $p_{jj}^{(n)} = 0$ for $n \neq kd_j$ and $p_{jj}^{(kd_j)} \to 0$ similarly as in (iii).

□

Remark. The relation

$$\overline{p}_{ij}^{(n)} = \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} \to \frac{f_{ij}}{\mu_j}, \quad i \neq j$$

holds also in the case that j is recurrent aperiodic.

Theorem 2.10. *A recurrent state j is null recurrent if and only if $p_{jj}^{(n)} \to 0$ as $n \to \infty$.*

Proof. Let j be a recurrent state such that $p_{jj}^{(n)} \to 0$. Then $\frac{1}{n}\sum_{k=1}^{n} p_{jj}^{(k)} \to 0$ and according to the Tauber theorem (Appendix A),

$$\lim_{s\to 1-} (1-s) P_{jj}(s) = 0,$$

which according to (2.16) implies

$$\lim_{s\to 1-} \frac{1-s}{1-F_{jj}(s)} = \frac{1}{F'_{jj}(1)} = \frac{1}{\mu_j} = 0,$$

thus j is null recurrent. The rest of the proof follows from Theorem 2.9.

□

2.4 Decomposition of the state space

Definition 2.7. We say that state j is *accessible* or *can be reached* from a state i, $i \to j$, if there exists $n \in \mathbb{N}_0$ such that $p_{ij}^{(n)} > 0$. If $p_{ij}^{(n)} = 0$ for all $n \in \mathbb{N}_0$, we say that j *is not accessible* from i. If j is accessible form i and i is accessible from j, we say that states i and j *communicate*, $i \leftrightarrow j$.

Remark. Every state is accessible from itself, since $p_{jj}^{(0)} = 1$ (since $p_{ij}^{(0)} = \delta_{ij}$).

Definition 2.8. A non-empty set of states $C \subset S$ is called to be *closed* if $p_{ij}^{(n)} = 0\ \forall n \in \mathbb{N}_0, \forall i \in C$ a $j \notin C$. The smallest closed set containing a set C is called *the closure* of the set C. A closed set of states is called to be *irreducible,* if it does not contain any closed proper subset.

Theorem 2.11. *A set $C \subset S$ of states is closed if and only if $p_{ij} = 0$ for all $i \in C,\ j \notin C$.*

Proof. 1. Let $p_{ij} = 0\ \forall\ i \in C,\ j \notin C$. Obviously, $p_{ij}^{(1)} = 0\ \forall\ i \in C,\ j \notin C$. Further, let us use the mathematical induction and assume that for an integer $n \geq 1$ the relation $p_{ij}^{(n)} = 0\ \forall\ i \in C,\ j \notin C$ is satisfied. According to the Chapman-Kolmogorov equation (2.8) we get

$$p_{ij}^{(n+1)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}.$$

If $k \in C$, then $p_{kj} = 0$ from the assumption of the theorem. If $k \notin C$, then $p_{ik}^{(n)} = 0$ by the induction assumption.

2. Let C be closed, then $\forall\ i \in C,\ j \notin C$, $p_{ij}^{(n)} = 0\ \forall\ n \in \mathbb{N}_0$, thus also for $n = 1$. □

Remark. When deleting rows and columns of a probability transition matrix $\boldsymbol{P}$ that correspond to states outside a closed subset $C \subset S$, we obtain again a stochastic matrix.

The one element-set $\{j\}$ is closed if $p_{jj} = 1$. The closure of the one-element set $\{j\}$ is the set of all states accessible from j including j.

Definition 2.9. State j is called *absorbing* if the set $\{j\}$ is closed, i.e., if $p_{jj} = 1$.

Definition 2.10. A Markov chain is called *irreducible* if the only closed subset is the set of all states (it means that all the states communicate each other). Otherwise it is *reducible.*

Example 2.11. In the gambler's ruin problem (Example 2.3) states $0, 1, \dots, a$, are accessible from states $1, 2, \dots, a-1$, from state 0 only 0 is accessible, similarly, from a only a is accessible.

Sets $\{0\}, \{a\}$ are closed, set $\{1, 2, \dots, a-1\}$ is not closed, since, e.g., state 0 is accessible from state 1. States 0 and a are absorbing, the chain is reducible (the state space obtains closed proper subsets $\{0\}, \{a\}$). In the car insurance model (Example 2.7) all the states are mutually accessible. The only closed set is the set of all states, the chain is irreducible.

Example 2.12. Consider the Markov chain with the state space $S = \{1, 2, 3, \dots\}$, initial distribution $\boldsymbol{p} = (1, 0, \dots)^T$ and with transition probabilities $p_{i,i+1} = 1,\ i \geq 1$. The state space contains closed subsets $\{i, i+1, \dots\}, i \geq 1$. The chain is reducible.

Theorem 2.12. *A Markov chain with finite number of states is reducible if and only if the transition probability matrix (after a permutation of rows and columns) is of the form*

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_1 & \mathbf{0} \\ \boldsymbol{A} & \boldsymbol{B} \end{pmatrix} \tag{2.24}$$

where $\boldsymbol{P}_1, \boldsymbol{B}$ are square matrices.

Proof. 1. Let $\boldsymbol{P}$ be of the form given in (2.24), where $\boldsymbol{P}_1, \boldsymbol{B}$ are square matrices. Then $\boldsymbol{P}_1$ is a stochastic matrix that corresponds a proper closed subset of states. The chain is reducible.

2. Let the chain be reducible, the state space contains a proper closed subset. We can change the ordering and assign the lowest orders to the states from this closed subset. Applying the same permutation to the rows and columns of the transition matrix we get $\boldsymbol{P}$ in the form given by (2.24). □

Example 2.13. Consider a chain with the transition matrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We can see that subset $\{1,5\}$ is closed. Permutation $(1,5,2,3,4)$ of rows and columns changes the matrix into

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \end{pmatrix}.$$

Remark. The assumption that $\boldsymbol{P}_1, \boldsymbol{B}$ are squared matrices is crucial. For example, the chain with the probability transition matrix

$$\boldsymbol{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}$$

is irreducible, because $p_{13}^{(2)} > 0$, thus 3 is accessible from 1 (after two steps).

Definition 2.11. We say that two states of a Markov chain are of the same type if they are either both transient or both positive recurrent or null recurrent, both either aperiodic or periodic with the same period.

Theorem 2.13. *If states i and j communicate, they are of the same type.*

Proof. Let $j$be accessible from i and i be accessible from j. Then there exist N and M such that $p_{ij}^{(N)} = \alpha > 0, p_{ji}^{(M)} = \beta > 0$. Using Chapman-Kolmogorov equation for any $n \geq 0$ gives us

$$p_{ii}^{(N+M+n)} = \sum_{k \in S} p_{ik}^{(N)} p_{ki}^{(M+n)} \geq p_{ij}^{(N)} p_{ji}^{(M+n)},$$

$$p_{ji}^{(M+n)} = \sum_{k \in S} p_{jk}^{(n)} p_{ki}^{(M)} \geq p_{jj}^{(n)} p_{ji}^{(M)},$$

thus

$$p_{ii}^{(N+M+n)} \geq \alpha\beta p_{jj}^{(n)} \tag{2.25}$$

and similarly,

$$p_{jj}^{(N+M+n)} \geq \alpha\beta p_{ii}^{(n)}. \tag{2.26}$$

If i is recurrent, then (2.26) and Theorem 2.5 imply that j is recurrent, if i is null recurrent, then according to (2.25) and Theorem 2.10 j is also null recurrent, if i is positive recurrent, then (2.26) implies that j is positive recurrent. If i is transient then j is also transient due to (2.25) and Theorem 2.5. The reasoning for states j a i je symmetric.

Now, let us assume that i and j communicate, let i be periodic with the period d_i. By (2.25) for $n = 0$ we have $p_{ii}^{(M+N)} \geq \alpha\beta > 0$, and $M + N$ is divided by d_i. Then, $p_{ii}^{(M+N+n)}$ equals zero for n not divided by d_i, from here and from (2.25) $p_{jj}^{(n)} = 0$ for n not divided by d_i, hence, the greatest common divisor of $\{n : p_{jj}^{(n)} > 0\} \geq d_i$, it means that j is periodic with the period $d_j \geq d_i$. If we exchange the role of i and j, we obtain $d_i \geq d_j$, thus $d_i = d_j$. □

Theorem 2.14. *In the irreducible Markov chain all the states are of the same type.*

Proof. It follows from the previous theorem. □

Theorem 2.15. *Let j be recurrent and k be accessible from j. Then*

(i) *k is recurrent,*

(ii) *j is accessible from k,*

(iii) $f_{jk} = P_j(\tau_k(1) < \infty) = P_k(\tau_j(1) < \infty) = f_{kj} = 1.$

Proof. If we prove (ii) we also prove (i), since then j and k will be states that communicate and are of the same type.

Let state k be accessible from j. Then there exists m such that $p_{jk}^{(m)} > 0$; let m be the smallest integer with this property.

Suppose that j is not accessible from k, then

$$P_k(X_n \neq j, \forall\, n \geq 1) = P_k(\tau_j(1) = \infty) = 1.$$

Since j is recurrent, by Theorem 2.8 $P_j(N_j = \infty) = 1$, and thus $P_j(X_m \neq j, X_{m+1} \neq j, \dots) = 0$ (m is finite). Then

$$\begin{aligned}
0 &= P_j(X_m \neq j, X_{m+1} \neq j, \dots) \geq P_j(X_m = k, X_{m+1} \neq j, \dots) \\
&= P(X_m = k, X_{m+1} \neq j, \dots | X_0 = j) \\
&= P(X_{m+1} \neq j, X_{m+2} \neq j, \dots | X_m = k, X_0 = j) P(X_m = k | X_0 = j) \\
&= P(X_{m+1} \neq j, X_{m+2} \neq j, \dots | X_m = k) P(X_m = k | X_0 = j) \\
&= P(X_1 \neq j, X_2 \neq j, \dots | X_0 = k) p_{jk}^{(m)} \\
&= P_k(X_n \neq j, \forall\, n \geq 1) p_{jk}^{(m)} = p_{jk}^{(m)} > 0,
\end{aligned}$$

which is a contradiction, thus, j is accessible from k and (ii) and (i) hold.

Similarly, by using conditioning, the Markov property and homogeneity, we have

$$\begin{aligned} 1 - f_{jj} &= P_j(\tau_j(1) = \infty) \geq P_j(\tau_j(1) = \infty, X_m = k) \\ &= P_j(X_1 \neq j, \dots, X_{m-1} \neq j, X_m = k, X_{m+1} \neq j, \dots) \\ &= p_{jk}^{(m)} P_k(\tau_j(1) = \infty) = p_{jk}^{(m)}[1 - f_{kj}]. \end{aligned}$$

Since j is recurrent, $f_{jj} = 1$ holds and $f_{kj} = 1$, too (because $p_{jk}^{(m)} > 0$). Since k is recurrent, $f_{jk} = 1$ which can be proved in the same way.

□

We can conclude that the set of states accessible from a recurrent state is closed and irreducible. The previous theorem enables us to decompose the state space S of a Markov chain in the following way. Let j_1 be the smallest order of a recurrent state, let C_1 be the set of all states accessible from j_1; let j_2 be the smallest order of a recurrent state among those that do not belong to C_1, let C_2 be the set of all states accessible from j_2 etc.

Proceeding in this way we realize that S is of the form

$$S = T \cup C_1 \cup C_2 \cup \dots,$$

where T is the set of transient states and $C_1, C_2, \dots$ are disjoint closed irreducible sets of recurrent states.

If S is finite, the transitition probability matrix is of the form

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_1 & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{P}_2 & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{P}_r & \mathbf{0} \\ \boldsymbol{Q}_1 & \boldsymbol{Q}_2 & \dots & \boldsymbol{Q}_r & \boldsymbol{Q}_{r+1} \end{pmatrix},$$

where $\boldsymbol{P}_1, \dots, \boldsymbol{P}_r$ are squared matrices of transition probabilities between recurrent states in subsets $C_1, \dots, C_r$ and $\boldsymbol{Q}_1, \dots, \boldsymbol{Q}_{r+1}$ are matrices of probabilities of transitions from transient states.

Theorem 2.16. *In the chain with a finite number of states not all states are transient.*

Proof. Suppose that all the states are transient. Then by (i) in Theorem 2.9 $p_{ij}^{(n)} \to 0$ as $n \to \infty \ \forall \ i, j \in S$. From here

$$\lim_{n\to\infty} \sum_{j\in S} p_{ij}^{(n)} = \sum_{j\in S} \lim_{n\to\infty} p_{ij}^{(n)} = 0, \quad \forall \ i \in S,$$

which is a contradiction since the matrix $\boldsymbol{P}^{(n)} = \{p_{ij}^{(n)}\}$ is stochastic and all the row totals must be equal to one. □

Example 2.14. In a chain with infinitely many states all the states can be transient. Let us consider the chain with states $S = \{1, 2, 3, \dots\}$, wit the initial distribution $\boldsymbol{p} = (1, 0, \dots)^T$ and the transition probabilities $p_{i,i+1} = 1, \ i \geq 1$. Then $p_{ii}^{(n)} = 0 \ \forall \ i \in S$ and $\forall \ n \geq 1$, thus $\sum p_{ii}^{(n)} < \infty, i \in S$. All the states are transient.

Theorem 2.17. *In the chain with a finite number of states there exist no null recurrent states.*

Proof. Let i be a null recurrent state, let C be the set of states accessible from i. According to Theorem 2.15 C is the closed irreducible set of null recurrent states with corresponding transition probability matrix $\boldsymbol{P}_C = \{\widetilde{p}_{ij}\}$. By (iv) in Theorem 2.9, $\widetilde{p}_{ij}^{(n)} \to 0 \ \forall i, j \in C$. But then

$$\lim_{n\to\infty} \sum_{j\in C} \widetilde{p}_{ij}^{(n)} = \sum_{j\in C} \lim_{n\to\infty} \widetilde{p}_{ij}^{(n)} = 0, \ i \in C,$$

which is a contradiction, since $\boldsymbol{P}_C$ and $\boldsymbol{P}_C^{(n)} = \{\widetilde{p}_{ij}^{(n)}\}$ are stochastic matrices. □

Theorem 2.18. *In an irreducible Markov chain with a finite number of states all the states are positive recurrent.*

Proof. A consequence of Theorems 2.14, 2.16 and 2.17. □

2.5 Absorption probabilities

In the previous section we have shown that the state space of a Markov chain with recurrent states can be decomposed into a disjoint union

$$S = T \cup C_1 \cup C_2 \cup \dots,$$

where T is the set of transient states and $C_1, C_2, \dots$ are closed irreducible sets of recurrent states. Sets C_j can consist of one element only; these are absorbing states j such that $p_{jj} = 1$.

Consider a chain $\{X_n\}$ with the set T of transient states and define random variable

$$\tau = \inf\{n \geq 0 : X_n \notin T\},$$

that denotes *the exit time from the set* T. Obviously, τ is the random variable taking values $0, 1, \dots,$ but let us remark that value $\tau = \infty$, can be achieved with non-zero probability, too, see Example 2.14, where $T = S$; in this case $P_i(\tau = \infty) = 1 \quad \forall\, i \in S$.

Theorem 2.19. *In a chain with finite number of states (shortly: in finite chain)*

$$P_i(\tau = \infty) = 0, \quad i \in T.$$

Proof. Since S and consequently T is finite, $[\tau = \infty] = \cup_{j \in T}[N_j = \infty]$ and

$$P_i(\tau = \infty) = P_i\left(\bigcup_{j \in T}[N_j = \infty]\right) \leq \sum_{j \in T} P_i(N_j = \infty) = 0,$$

since according to Theorem 2.8 for a transient state j, $P_i(N_j = \infty) = 0$ for all i. (Recall that N_j denotes the number of visits to j after leaving the initial state.) □

In the next we will always assume that $P_i(\tau < \infty) = 1$ for all

Let X_τ denote the state into which the chain enters when leaving the set of transient states T, ,i.e., at random time τ. Define probabilities

$$u_{ij} = P_i(X_\tau = j), \quad i \in T, j \in T^c. \tag{2.27}$$

When j is absorbing, u_{ij} means the probability that the chain starting from transient state i, is absorbed at state j. If C_k is a closed irreducible set of recurrent states then the probability that the chain starting from i is absorbed in C_k is

$$u_i(C_k) = P_i(X_\tau \in C_k) = \sum_{j \in C_k} u_{ij}. \tag{2.28}$$

Let us show how to compute probabilities u_{ij} by using transition probabilities p_{kl}.

Theorem 2.20. *Probabilities u_{ij} defined by (2.27) satisfy*

$$u_{ij} = p_{ij} + \sum_{\nu \in T} p_{i\nu} u_{\nu j}, \quad i \in T, j \in T^c. \tag{2.29}$$

Proof. Denote by

$$u_{ij}^{(n)} = P_i(X_\tau = j, \tau = n), \quad i \in T, j \in T^c$$

the probability that the chain exits the set of transient states and hits the recurrent state j just in time n. Since

$$[X_\tau = j] = \bigcup_{n=0}^{\infty} [X_\tau = j, \tau = n],$$

we have

$$u_{ij} = \sum_{n=0}^{\infty} u_{ij}^{(n)}, \quad i \in T, j \in T^c.$$

Obviously, probabilities $u_{ij}^{(n)}$ satisfy

$$\begin{aligned} u_{ij}^{(0)} &= 0, \\ u_{ij}^{(1)} &= p_{ij}, \\ u_{ij}^{(n)} &= \sum_{\nu \in T} p_{i\nu} u_{\nu j}^{(n-1)}, \quad n \geq 2. \end{aligned}$$

Notice that for $n \geq 2$, the relation for $u_{ij}^{(n)}$ follows easily from

$$u_{ij}^{(n)} = P_i(X_1 \in T, \ldots, X_{n-1} \in T, X_n = j \in T^c)$$

by conditioning on $[X_1 = \nu \in T, X_0 = i]$ and using the Markov property. We get

$$u_{ij} = \sum_{n=1}^{\infty} u_{ij}^{(n)} = p_{ij} + \sum_{n=2}^{\infty} \sum_{\nu \in T} p_{i\nu} u_{\nu j}^{(n-1)} = p_{ij} + \sum_{\nu \in T} p_{i\nu} \sum_{n=2}^{\infty} u_{\nu j}^{(n-1)} = p_{ij} + \sum_{\nu \in T} p_{i\nu} u_{\nu j}.$$

□

Notice that the transition probability matrix can be written in the form

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}^* & \boldsymbol{0} \\ \boldsymbol{Q} & \boldsymbol{R} \end{pmatrix}, \tag{2.30}$$

where $\boldsymbol{P}^* = \{p_{ij}, i, j \in T^c\}$, $\boldsymbol{Q} = \{p_{ij}, i \in T, j \in T^c\}$, $\boldsymbol{R} = \{p_{ij}, i, j \in T\}$.

Put $\boldsymbol{U} = \{u_{ij}, i \in T, j \in T^c\}$. Then (2.29) becomes

$$\boldsymbol{U} = \boldsymbol{Q} + \boldsymbol{R}\boldsymbol{U}. \tag{2.31}$$

For finite dimensional matrices we get $(\boldsymbol{I} - \boldsymbol{R})\boldsymbol{U} = \boldsymbol{Q}$, where $\boldsymbol{I}$ is the identity matrix of the same order as matrix $\boldsymbol{R}$. If an inverse matrix to $\boldsymbol{I} - \boldsymbol{R}$ does exist then there is the only solution

$$\boldsymbol{U} = (\boldsymbol{I} - \boldsymbol{R})^{-1}\boldsymbol{Q}. \tag{2.32}$$

The existence of the inverse matrix to $\boldsymbol{I} - \boldsymbol{R}$ follows from the following assertion.

Theorem 2.21. *Consider a Markov chain with transition probability matrix (2.30). Let the set T of transient states be finite. Then the matrix $\boldsymbol{I} - \boldsymbol{R}$ is regular and*

$$(\boldsymbol{I} - \boldsymbol{R})^{-1} = \sum_{k=0}^{\infty} \boldsymbol{R}^k.$$

Proof. If T is finite, $\boldsymbol{R}$ is finite square matrix of probability transitions between transient states only. It follows from (2.30) that $\boldsymbol{R}^n = \{p_{ij}^{(n)}, i, j \in T\}$. By Theorem 2.9 $p_{ij}^{(n)} \to 0$ při $n \to \infty$ $\forall\, i, j \in T$, thus $\boldsymbol{R}^n$ converges to the null matrix as n is increasing. The rest of the proof follows from Theorem B.2 in Appendix B.

□

Remark. Matrix $\boldsymbol{F} = (\boldsymbol{I} - \boldsymbol{R})^{-1}$ is called the *fundamental matrix* of a Markov chain.

Example 2.15. A university course consists of 5 stages (study units). At the end of each study stage a student's progress is classified as it follows: he successfully passes exams and moves to the next stage (with probability p), fails and repeats the same stage (with probability r) and leaves the study without graduation with probability q, $p+q+r=1$. We assume that probabilities p, q, r are constant and do not depend on the previous performances of the student. Then the education process can be described by a Markov chain with states 1- leaving the study, 2 - successfully graduating from the study, 3 - study in the first stage, 2 - in the second stage, ..., 7 - study the fifth stage. The probability transition matrix is

$$\boldsymbol{P} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ q & 0 & r & p & 0 & 0 & 0 \\ q & 0 & 0 & r & p & 0 & 0 \\ q & 0 & 0 & 0 & r & p & 0 \\ q & 0 & 0 & 0 & 0 & r & p \\ q & p & 0 & 0 & 0 & 0 & r \end{pmatrix},$$

states "leaving the study" a "successfully graduating from the study" are absorbing, other states are transient. In this case

$$\boldsymbol{Q} = \begin{pmatrix} q & 0 \\ q & 0 \\ q & 0 \\ q & 0 \\ q & p \end{pmatrix}, \quad \boldsymbol{R} = \begin{pmatrix} r & p & 0 & 0 & 0 \\ 0 & r & p & 0 & 0 \\ 0 & 0 & r & p & 0 \\ 0 & 0 & 0 & r & p \\ 0 & 0 & 0 & 0 & r \end{pmatrix}$$

and

$$\boldsymbol{F} = \begin{pmatrix} \frac{1}{p+q} & \frac{p}{(p+q)^2} & \frac{p^2}{(p+q)^3} & \frac{p^3}{(p+q)^4} & \frac{p^4}{(p+q)^5} \\ 0 & \frac{1}{p+q} & \frac{p}{(p+q)^2} & \frac{p^2}{(p+q)^3} & \frac{p^3}{(p+q)^4} \\ 0 & 0 & \frac{1}{p+q} & \frac{p}{(p+q)^2} & \frac{p^2}{(p+q)^3} \\ 0 & 0 & 0 & \frac{1}{p+q} & \frac{p}{(p+q)^2} \\ 0 & 0 & 0 & 0 & \frac{1}{p+q} \end{pmatrix}.$$

A student studying the first stage leaves study with probability

$$u_{31} = q\left(\frac{1}{p+q} + \frac{p}{(p+q)^2} + \frac{p^2}{(p+q)^3} + \frac{p^3}{(p+q)^4} + \frac{p^4}{(p+q)^5}\right)$$

and successfully graduates with probability

$$u_{32} = \frac{p^5}{(p+q)^5} = 1 - u_{31}.$$

Example 2.16. Let us determine the ruin probability for the player A from Example 2.3. We know that states $0, a$ are absorbing, other states are transient. We want to find the probability of absorption from the transient state $i, 1 \le i \le a-1$ (starting capital of player A) to state 0. According to (2.29) absorbing probabilities have to satisfy

$$u_{i0} = p_{i0} + \sum_{\nu=1}^{a-1} p_{i\nu} u_{\nu 0}, \; i = 1, \ldots, a-1,$$

in our case,

$$\begin{aligned} u_1 - q - pu_2 &= 0, \\ u_i - qu_{i-1} - pu_{i+1} &= 0, \quad i = 2, \ldots, a-2, \\ u_{a-1} - qu_{a-2} &= 0, \end{aligned} \tag{2.33}$$

where for simplicity u_{i0} is denoted by u_i. Put

$$u_0 = 1, \quad u_a = 0. \tag{2.34}$$

Then the system (2.33) satisfies the homogeneous difference equation

$$-pu_{i+1} + u_i - qu_{i-1} = 0 \tag{2.35}$$

with marginal conditions (2.34). The characteristic polynomial of this equation is $-p\lambda^2 + \lambda - q$, with the roots 1 and $\frac{q}{p}$.

A general solution to the difference equation (2.35) is

$$\begin{aligned} u_i &= c_1 + c_2 \left(\frac{q}{p}\right)^i, \quad && p \neq q \\ u_i &= c_1 + ic_2, && p = q. \end{aligned}$$

From the marginal conditions and for $p \neq q$ we get

$$c_1 = -\frac{\left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^a}, \quad c_2 = \frac{1}{1 - \left(\frac{q}{p}\right)^a},$$

thus,

$$u_i = \frac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^a}, \quad i = 1, \ldots, a-1.$$

For $p = q$ we similarly have

$$c_1 = 1, \quad c_2 = -\frac{1}{a},$$

and therefore

$$u_i = 1 - \frac{i}{a}, \quad i = 1, \ldots a-1.$$

Consider now random variable W_j that denotes the total time the chain spends in a transient state j. Obviously,

$$W_j = \begin{cases} N_j, & X_0 = i \neq j, \\ N_j + 1, & X_0 = j. \end{cases}$$

Then the mean time spent in j is

$$E_iW_j = E_iN_j = E_i\left(\sum_{n=1}^{\infty} I(X_n = j)\right) = \sum_{n=1}^{\infty} P_i(X_n = j) = \sum_{n=1}^{\infty} p_{ij}^{(n)}, \quad i, j \in T, i \neq j,$$

$$E_jW_j = E_jN_j + 1 = \sum_{n=1}^{\infty} p_{jj}^{(n)} + 1, \quad j \in T,$$

and since $p_{ij}^{(0)} = \delta_{ij}$, we have

$$E_iW_j = \sum_{n=0}^{\infty} p_{ij}^{(n)}, \quad i, j \in T.$$

If T is finite then by using Theorem 2.21

$$E_iW_j = \varphi_{ij}, \quad i, j \in T,$$

where φ_{ij} are the corresponding elements of the fundamental matrix $\boldsymbol{F} = (\boldsymbol{I} - \boldsymbol{R})^{-1}$.

Denote by $W = \sum_{j \in T} W_j$ the total time spent in the set T of transient states. Then

$$E_iW = \sum_{j \in T} E_iW_j = \boldsymbol{F}_i\mathbf{1},$$

where $\boldsymbol{F}_i$ is the row of $\boldsymbol{F}$ that corresponds to the initial state i and $\mathbf{1} = (1, 1, \ldots, 1)^T$.

Up to now we know that if S (and thus T) is finite, then $P_i(\tau = \infty) = 0, \ i \in T$ and the system of equations (2.29), (2.31), respectively, has only one solution $\boldsymbol{U} = \boldsymbol{FQ}$, where $\boldsymbol{F}$ is the fundamental matrix. In case that S is not finite we must admit that(see Example 2.14) that $P_i(\tau = \infty) > 0$ and a question of a unique solution of (2.29) arises.

Theorem 2.22. *The system (2.29) has a unique solution* $0 \leq u_{ij} \leq 1, \ i \in T, j \in T^c$ *if and only if the system*

$$x_i = \sum_{j \in T} p_{ij}x_j, \quad i \in T \tag{2.36}$$

has the solution $x_i = 0 \ \forall i \in T$*, only. This condition is equivalent to condition*

$$P_i(\tau = \infty) = 0, \quad i \in T. \tag{2.37}$$

Proof. See Resnick (1992), Proposition 2.11.1; here we show only that (2.36) and (2.37) are equivalent.

Denote $v_i = P_i(\tau = \infty),\ i \in T$. We will show that v_i solves (2.36). Denote $v_i^{(n)} = P_i(\tau > n)$ the probability that at time n the chain is still at the set T. Then

$$\begin{aligned} v_i^{(0)} &= P_i(\tau > 0) = 1, \quad i \in T, \\ v_i^{(1)} &= P_i(\tau > 1) = P_i(X_1 \in T) = \sum_{\nu \in T} p_{i\nu}, \\ v_i^{(n+1)} &= P_i(x_1 \in T, \dots, X_{n+1} \in T) = \sum_{\nu \in T} p_{i\nu} v_\nu^{(n)}, \quad n \geq 0. \end{aligned}$$

Since events $A_n = [\tau > n]$ satisfy $A_{n+1} \subset A_n \subset \dots$, it holds $v_i = \lim_{n\to\infty} v_i^{(n)}$ and thus

$$v_i = \lim_{n\to\infty} \sum_{\nu \in T} p_{i\nu} v_\nu^{(n)} = \sum_{\nu \in T} p_{i\nu} v_\nu$$

(we can do it since $0 \leq p_{i\nu} v_\nu \leq p_{i\nu}$ and $\sum_{\nu \in T} p_{i\nu} \leq 1$.)

Let $0 \leq \widetilde{v}_i \leq 1$ be other solution to (2.36). We will show that $\widetilde{v}_i \leq v_i, i \in T$. It holds

$$\widetilde{v}_i = \sum_{\nu \in T} p_{i\nu} \widetilde{v}_\nu \leq \sum_{\nu \in T} p_{i\nu} = v_i^{(1)},$$

and further, by induction for n we get

$$\widetilde{v}_i \leq v_i^{(n)} \text{ for all } n,$$

since if $\widetilde{v}_i \leq v_i^{(k)}$ for some k, then by using the induction

$$\widetilde{v}_i = \sum_{\nu \in T} p_{i\nu} \widetilde{v}_\nu \leq \sum_{\nu \in T} p_{i\nu} v_\nu^{(k)} = v_i^{(k+1)}.$$

From here

$$\widetilde{v}_i \leq \lim_{n\to\infty} v_i^{(n)} = v_i, \quad i \in T.$$

So, if $v_i = 0\ \forall i \in T$, it is the only solution to (2.36) in the interval $[0, 1]$, and, on the contrary, if there exists only the trivial solution to (2.36) then $v_i = 0\ \forall i \in T$. □

In Example 2.14 we have considered a chain with a countable many transient states. The following theorem help us to decide about states of a Markov chain with infinitely many states.

Theorem 2.23. *In an irreducible Markov chain with the state space* $S = \{0, 1, \ldots, \}$ *all the states are recurrent if and only if the only solution of the system of equations*

$$x_i = \sum_{j=1}^{\infty} p_{ij} x_j, \quad i = 1, 2 \ldots \tag{2.38}$$

in the interval $[0, 1]$ *is the trivial solution* $x_i = 0, i = 1, 2 \ldots$*. All the states are transient if and only if (2.38) has a non-trivial solution in* $[0, 1]$.

Proof. Let us consider the subset $T = \{1, 2, \ldots\}$ and the decomposition $S = \{0\} \cup T$. Then,

$$\tau = \inf\{n \geq 0 : X_n \notin T\} = \inf\{n \geq 0 : X_n = 0\}$$

is the exit time from T and similarly as in Theorem 2.22 it can be proved that $x_i = 0$ is the unique solution of the system (2.38) in the interval $[0, 1]$ if and only if (2.37) holds true, i.e., if $P_i(\tau = \infty) = 0, i = 1, 2, \ldots$. Notice that in the presented proof of Theorem 2.22 we have not used the assumption that T is a set of transient states. For $i \neq 0$ we have

$$P_i(\tau = \infty) = 1 - P_i(\tau < \infty) = 1 - f_{i0}.$$

Let all the states are recurrent; then according to Theorem 2.15 $f_{i0} = 1$ for all $i = 1, 2, \ldots$, thus (2.37) and consequently (2.38) hold. On the other hand, let the only solution to (2.38) in $[0, 1]$ is the trivial one. Then (2.37) holds and consequently $f_{i0} = 1, i = 1, 2 \ldots$. Then,

$$f_{00} = P_0(\tau_0(1) < \infty) = p_{00} + \sum_{i=1}^{\infty} p_{0i} f_{i0} = \sum_{i=0}^{\infty} p_{0i} = 1,$$

which means that 0 is the recurrent state. Since the chain is irreducible, all the states are recurrent.

□

2.6 Stationary distribution

Definition 2.12. Let $\{X_n, n \in \mathbb{N}_0\}$ be a homogeneous Markov chain with a state space S and a transition probability matrix $\boldsymbol{P}$. Let $\boldsymbol{\pi} = \{\pi_j, j \in S\}$ be a probability distribution on S, i.e.,. $\pi_j \geq 0, j \in S$, $\sum_{j \in S} \pi_j = 1$. Then $\boldsymbol{\pi}$ is called *a stationary distribution* of the chain, if

$$\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \boldsymbol{P}, \tag{2.39}$$

for column vectors, or, component-wise,

$$\pi_j = \sum_{k \in S} \pi_k p_{kj}, \quad j \in S.$$

Theorem 2.24. *Suppose that the initial distribution of a homogeneous Markov chain is stationary in the sense of (2.39). Then the Markov chain is strictly stationary stochastic process, i.e., for all $n \in \mathbb{N}_0, k \in \mathbb{N}_0$ and any $i_0, \ldots, i_n \in S$*

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = P(X_k = i_0, X_{k+1} = i_1, \ldots, X_{k+n} = i_n).$$

Especially, for absolute probabilities

$$p_j(n) = P(X_n = j) = \pi_j, \quad j \in S,$$

where π_j are initial stationary probabilities.

Proof. Let $\boldsymbol{\pi}$ be a stationary distribution. First, we prove by induction that for every $n \geq 1$

$$\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \boldsymbol{P}^n, \tag{2.40}$$

i.e.,

$$\pi_j = \sum_{k \in S} \pi_k p_{kj}^{(n)}, \quad j \in S.$$

For $n = 1$ this follows from the definition and if it holds for some $n \geq 1$, then

$$\boldsymbol{\pi}^T \boldsymbol{P}^{n+1} = \boldsymbol{\pi}^T \boldsymbol{P}^n \boldsymbol{P} = \boldsymbol{\pi}^T \boldsymbol{P} = \boldsymbol{\pi}^T.$$

Let $\boldsymbol{\pi}$ be the initial distribution. From (2.9) we get

$$\boldsymbol{p}(n)^T = \boldsymbol{\pi}^T \boldsymbol{P}^n = \boldsymbol{\pi}^T.$$

By using Theorems 2.1 and (2.40) we get

$$\begin{gathered} P(X_k = i_0, X_{k+1} = i_1, \ldots, X_{k+n} = i_n) = \\ \sum_{j \in S} P(X_0 = j, X_k = i_0, \ldots, X_{k+n} = i_n) = \sum_{j \in S} \pi_j p_{j i_0}^{(k)} p_{i_0 i_1} \ldots p_{i_{n-1} i_n} \\ = \pi_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n} = P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n). \end{gathered}$$

□

Theorem 2.25. *Let a Markov chain be irreducible. Then the following holds true:*

(i) *If all the states are transient or all are null recurrent, the stationary distribution does not exist.*

(ii) *If all the states are positive recurrent there exists the unique stationary distribution. If moreover all the states are aperiodic, then the stationary probabilities π_j satisfy*

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n)} > 0, \quad i,j \in S$$

and

$$\pi_j = \lim_{n\to\infty} p_j(n) > 0, \quad j \in S.$$

When all the states are periodic then

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_{ij}^{(k)} > 0, \quad i,j \in S,$$

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} p_j(k) > 0, \quad j \in S.$$

Proof. (i) If all the states are transient or null recurrent then according to Theorem 2.9 $p_{ij}^{(n)} \to 0$ as $n \to \infty$ $\forall i,j \in S$. Let us suppose that the stationary distribution does exist. Then by (2.40) $\pi_j = \sum_{i\in S} \pi_i p_{ij}^{(n)}$, and we get for all $j \in S$

$$\pi_j = \lim_{n\to\infty} \sum_{i\in S} \pi_i p_{ij}^{(n)} = \sum_{i\in S} \pi_i \lim_{n\to\infty} p_{ij}^{(n)} = 0,$$

which is a contradiction since in this case $\{\pi_j\}$ is not the probability distribution.

(ii) If all the states are positive recurrent then by Theorem 2.15 $f_{ij} = 1, \forall i,j \in S$. If they are aperiodic, then according to Theorem 2.9 (ii), $p_{ij}^{(n)} \to \frac{1}{\mu_j} \forall i,j \in S$, where $\mu_j > 0,\ j \in S$. If all the states are periodic with the same period then by Theorem 2.9 (iii), $\overline{p}_{ij}^{(n)} = \frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)} \to \frac{1}{\mu_j}$ for $i \neq j$. By using the same assertion we can easily prove that the same result holds true also for $\overline{p}_{jj}^{(n)}$. Further, it can be proved that (2.40) holds with $p_{ij}^{(n)}$ replaced by $\overline{p}_{ij}^{(n)}$. Thus, in the proof we can confine ourselves to aperiodic states.

Now, we will show that the stationary distribution does exist. Consider probabilities

$$p_{kj}^{(n+1)} = \sum_{i\in S} p_{ki}^{(n)} p_{ij}, \quad k,j \in S.$$

Is S is finite, we can apply the limit as $n \to \infty$ to ger

$$\frac{1}{\mu_j} = \sum_{i \in S} \frac{1}{\mu_i} p_{ij}, \quad j \in S. \tag{2.41}$$

If S is infinitely countable, then

$$p_{kj}^{(n+1)} = \sum_{i=0}^{\infty} p_{ki}^{(n)} p_{ij} \geq \sum_{i=0}^{N} p_{ki}^{(n)} p_{ij}, \quad N > 0,$$

and from here, for N fixed, using the limit as $n \to \infty$ we obtain

$$\frac{1}{\mu_j} \geq \sum_{i=0}^{N} \frac{1}{\mu_i} p_{ij}$$

and then, as $N \to \infty$

$$\frac{1}{\mu_j} \geq \sum_{i=0}^{\infty} \frac{1}{\mu_i} p_{ij}, \quad j \in S.$$

Notice that if

$$\frac{1}{\mu_j} > \sum_{i=0}^{\infty} \frac{1}{\mu_i} p_{ij},$$

for some j then

$$\sum_{j=0}^{\infty} \frac{1}{\mu_j} > \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \frac{1}{\mu_i} p_{ij} = \sum_{i=0}^{\infty} \frac{1}{\mu_i} \sum_{j=0}^{\infty} p_{ij} = \sum_{i=0}^{\infty} \frac{1}{\mu_i},$$

which is a contradiction. Thus (2.41) must hold. Since

$$\sum_{j=0}^{N} p_{kj}^{(n)} \leq \sum_{j \in S} p_{kj}^{(n)} = 1,$$

we have $\sum_{j \in S} \frac{1}{\mu_j} \leq 1$. The vector $\boldsymbol{\pi} = \{\pi_j, j \in S\}$, where

$$\pi_j = \frac{1}{\mu_j} \left(\sum_{i \in S} \frac{1}{\mu_i} \right)^{-1}, \quad j \in S,$$

satisfies (2.39), and thu, it is a stationary distribution.

Since $\boldsymbol{\pi}$ is the stationary distribution,

$$\pi_j = \sum_{i \in S} \pi_i p_{ij}^{(n)}, \quad i, j \in S,$$

and from here, using the limit as $n \to \infty$, we get

$$\pi_j = \sum_{i \in S} \pi_i \lim_{n\to\infty} p_{ij}^{(n)} = \sum_{i \in S} \pi_i \frac{1}{\mu_j} = \frac{1}{\mu_j}$$

(hence, $\sum_{j \in S} \frac{1}{\mu_j} = 1$).

Let $\{v_j, j \in S\}$ be another stationary distribution. Then for all $j \in S$ we have

$$v_j = \sum_{i \in S} v_i p_{ij} = \sum_{i \in S} v_i p_{ij}^{(n)}$$

and limiting this we get $v_j = \sum_{i \in S} v_i \frac{1}{\mu_j} = \frac{1}{\mu_j}$. Thus, there exists the only one stationary distribution and it holds

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_j} > 0, \quad i, j \in S.$$

For absolute probabilities we can use (2.9) and get

$$\lim_{n\to\infty} p_j(n) = \lim_{n\to\infty} \sum_{i \in S} p_i(0) p_{ij}^{(n)} = \sum_{i \in S} p_i(0) \frac{1}{\mu_j} = \frac{1}{\mu_j} = \pi_j, \quad j \in S.$$

□

Remark. The relation $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j,\ i, j \in S$ can be written as

$$\lim_{n\to\infty} \boldsymbol{P}^n = \boldsymbol{\Pi},$$

where

$$\boldsymbol{\Pi} = \begin{pmatrix} \pi_1 & \pi_2 & \cdots \\ \pi_1 & \pi_2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} \boldsymbol{\pi}^T \\ \boldsymbol{\pi}^T \\ \vdots \end{pmatrix}.$$

Theorem 2.26. *In an irreducible finite chain the stationary distribution always exists.*

Proof. A consequence of Theorems 2.18 and 2.25. □

Example 2.17. In the car insurance model (Example 2.7), $\{X_n\}$ is the Markov chain with states $0, 1, 2,$, the initial distribution $\boldsymbol{p} = (1, 0, 0)^T$ and with the transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} 1-a_0 & a_0 & 0 \\ 1-a_0 & 0 & a_0 \\ 1-a_0-a_1 & a_1 & a_0 \end{pmatrix},$$

where $a_0 = e^{-\lambda},\ a_1 = \lambda e^{-\lambda}$.

The chain is irreducible, all the states are positive recurrent and aperiodic. The stationary distribution exists and is given by (2.39), i.e., by the solution of the system of equations

$$\begin{aligned} \pi_0 &= \pi_0(1-a_0) + \pi_1(1-a_0) + \pi_2(1-a_0-a_1), \\ \pi_1 &= \pi_0 a_0 + \pi_2 a_1, \\ \pi_2 &= \pi_1 a_0 + \pi_2 a_0, \end{aligned}$$

which results to

$$\begin{aligned} \pi_0 &= \frac{1-a_0-a_0a_1}{1-a_0a_1} = \frac{1-e^{-\lambda}-\lambda e^{-2\lambda}}{1-\lambda e^{-2\lambda}}, \\ \pi_1 &= \frac{a_0(1-a_0)}{1-a_0a_1} = \frac{e^{-\lambda}(1-e^{-\lambda})}{1-\lambda e^{-2\lambda}}, \\ \pi_2 &= \frac{a_0^2}{1-a_0a_1} = \frac{e^{-2\lambda}}{1-\lambda e^{-2\lambda}}. \end{aligned}$$

Example 2.18. The existence of the stationary distribution can be used as a criterion for the state classification.

Let us consider The Markov chain with the state space $S = \{0, 1, \dots\}$ and the transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} q & p & 0 & 0 & \dots \\ q & 0 & p & 0 & \dots \\ 0 & q & 0 & p & \dots \\ & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

where $0 < p < 1, q = 1 - p$. The chain is irreducible, all the states are of the same type.

Search for a probability solution of the system of equations

$$\begin{aligned}
\pi_0 &= q\pi_0 + q\pi_1, \\
\pi_1 &= p\pi_0 + q\pi_2, \\
&\vdots \\
\pi_j &= p\pi_{j-1} + q\pi_{j+1}, \\
&\vdots
\end{aligned}$$

We get recursively

$$\pi_j = \left(\frac{p}{q}\right)^j \pi_0, \quad j = 0, 1, \ldots, \ \pi_0 > 0.$$

If $\frac{p}{q} < 1$ we get

$$\pi_0 = 1 - \frac{p}{q}, \quad \pi_j = (1 - \frac{p}{q}) \left(\frac{p}{q}\right)^j, \quad j \geq 1.$$

For $p < q$ the stationary distribution exists and thus all the states are positive recurrent. For $p \geq q$ the stationary distribution does not exist and all the states are either null recurrent or transient. We can further decide according to Theorem 2.23. The system (2.38) is of the form

$$\begin{aligned}
x_1 &= px_2, \\
x_i &= qx_{i-1} + px_{i+1}, \quad i \geq 2.
\end{aligned}$$

which is the system of difference equations

$$px_{i+1} - x_i + qx_{i-1} = 0, \quad i \geq 1$$

with the marginal condition $x_0 = 0$. Proceeding as in Example 2.16, we get the general solution

$$\begin{aligned}
x_i &= c_1 + c_2 \left(\frac{q}{p}\right)^i, & p \neq q, \\
x_i &= c_1 + ic_2, & p = q.
\end{aligned}$$

With the marginal condition $x_0 = 0$ we have

$$\begin{aligned}
x_i &= A\Big(1 - (\frac{q}{p})^i\Big), & p > q, \\
x_i &= Bi, & p = q,
\end{aligned}$$

where A, B are constants. We can see that for $p > q$ there exists a nontrivial solution of (2.38) in the interval $[0, 1]$, thus, all the states are transient; for $p = q$ there is only the trivial solution of (2.38) in $[0, 1]$, and all the states are null recurrent.

Example 2.19. The chain with the state space $S = \{0, 1, \dots\}$ and the transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & \dots \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & \dots \\ \frac{1}{4} & 0 & 0 & \frac{3}{4} & \dots \\ & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

is irreducible but the stationary distribution does not exist. The system (2.39) is of the form

$$\pi_0 = \frac{1}{2}\pi_0 + \frac{1}{3}\pi_1 + \frac{1}{4}\pi_2 + \dots,$$
$$\pi_j = \frac{j}{j+1}\pi_{j-1} = \frac{1}{j+1}\pi_0, \quad j = 1, 2 \dots$$

Since the sum $\sum_{j=0}^{\infty} \pi_j = \pi_0 \sum_{j=0}^{\infty} \frac{1}{j+1}$ is not convergent for any $\pi_0 > 0$, there is no positive solution to (2.39). Since the chain is irreducible, all the states are either transient or null recurrent. Again, we can decide by using Theorem 2.23. The system (2.38) is

$$x_j = \frac{j+1}{j+2} x_{j+1}, \quad j = 1, 2 \dots,$$

form here

$$x_j = \frac{j+1}{2} x_1, \quad j = 1, 2 \dots$$

and the only solution in $[0, 1]$ is $x_j = 0, \ \forall j$. All the states are null recurrent (and aperiodic since $p_{00} > 0$).

2.7 Limit theorems for frequencies of returns

Consider an irreducible Markov chain with positive recurrent states. We know from the proof of Theorem 2.25 that there exists the unique stationary distribution $\{\pi_j, j \in S\}$ and

$$\pi_j = \frac{1}{\mu_j} > 0, \quad j \in S,$$

where $\mu_j = E_j \tau_j(1) = E_j T_1$ is the mean value of the first return to state j.

Let $N_j(n)$ be the random variable that denotes the number of hittings of state j in the first n steps, i.e.,

$$N_j(n) = \sum_{\nu=1}^{n} I(X_\nu = j).$$

Then for $k = 1, 2 \ldots$

$$[N_j(n) < k] \iff [\tau_j(k) > n],$$

where $\tau_j(k)$ is time of the $k - th$ visit to j. Since the chain is irreducible and all states are positive recurrent, we have $P_i(\tau_j(k) < \infty) = 1 \ \forall i, j \in S, \ \forall k = 1, 2 \ldots.$

Now we can prove an analogy of the strong law of large numbers for relative frequencies.

Theorem 2.27. *In an irreducible Markov chain with positive recurrent states it holds*

$$\frac{N_j(n)}{n} \to \pi_j \qquad \text{as } n \to \infty \text{ almost surely}$$

for every $j \in S$.

Proof. Obviously, the following relations between random events hold:

$$\left[\frac{N_j(n)}{n} - \pi_j < \varepsilon\right] = [N_j(n) < n(\pi_j + \varepsilon)] \supseteq [N_j(n) < Z],$$

where Z denotes the integer part of $n(\pi_j + \varepsilon)$, and further,

$$[N_j(n) < Z] = [\tau_j(Z) > n] = \left[\frac{\tau_j(Z)}{Z} > \frac{n}{Z}\right] \supseteq \left[\frac{\tau_j(Z)}{Z} > \frac{n}{n(\pi_j + \frac{\varepsilon}{2})}\right]$$
$$= \left[\frac{\tau_j(Z)}{Z} - \frac{1}{\pi_j} > -\frac{\frac{\varepsilon}{2}}{\pi_j(\pi_j + \frac{\varepsilon}{2})}\right].$$

We know that $\tau_j(Z) = \sum_{k=1}^{Z} T_k$, where $T_1, T_2, \ldots, T_Z$ are independent random variables (see Theorem 2.6); if the initial state is j, they are equally distributed with the finite mean value $\mu_j = \frac{1}{\pi_j}$ and satisfy the strong law of large numbers, i.e.,

$$\frac{\tau_j(Z)}{Z} \to \frac{1}{\pi_j} \qquad \text{as } Z \to \infty \qquad \text{with probability 1} \tag{2.42}$$

(see, e.g., Štěpán, 1987, Theorem IV.2.2).

If the initial state differs from j, then $T_2, \ldots, T_Z$, are equally distributed and T_1 is finite with probability 1. From here we again get (2.42).

Thus, $\forall\ \varepsilon > 0,\ \forall\ \delta > 0\ \exists\ Z_0 = Z_0(\varepsilon, \delta)$ positive integer such that,

$$P\left(\frac{\tau_j(Z)}{Z} - \frac{1}{\pi_j} > -\varepsilon \quad \forall\ Z > Z_0\right) > 1 - \delta.$$

Since $Z \to \infty \Longleftrightarrow n \to \infty$, it holds that $\forall\, \varepsilon > 0,\ \forall\, \delta > 0\ \exists\, n_0 = n_0(\varepsilon, \delta)$ such that

$$P\left(\frac{N_j(n)}{n} - \pi_j < \varepsilon \quad \forall\, n > n_0\right) > 1 - \delta.$$

Analogously, we get $\forall\, \varepsilon > 0,\ \forall\, \delta > 0\ \exists\, n_0 = n_0(\varepsilon, \delta)$ such that

$$P\left(\frac{N_j(n)}{n} - \pi_j > -\varepsilon \quad \forall\, n > n_0\right) > 1 - \delta,$$

from which the assertion of the theorem follows.

□

Consider again an irreducible Markov chain with positive recurrent states (which implies that μ_j is finite for every j) and suppose that the variances $\sigma_j^2 = \operatorname{var}_j(\tau_j(1))$ are also finite. The we can state the following analogy of the central limit theorem.

Theorem 2.28. *In an irreducible Markov chain with positive recurrent states and finite variances of the return times it holds*

$$\lim_{n\to\infty} P\left(\frac{N_j(n) - \frac{n}{\mu_j}}{\sqrt{n\sigma_j^2/\mu_j^3}} \le x\right) = \Phi(x), \tag{2.43}$$

where Φ is the distribution function of the standard normal distribution.

Theorem 2.28 says , that for large n, random variable $N_j(n)$ is asymptotically normal with the mean

$$EN_j(n) \simeq n\frac{1}{\mu_j}$$

and the variance

$$\text{var } N_j(n) \simeq n\frac{\sigma_j^2}{\mu_j^3}.$$

Proof. We will simplify the proof assuming that the initial state is j. Since $N_j(n)$ is an integer valued random variable, it holds

$$P\left(\frac{N_j(n) - \frac{n}{\mu_j}}{\sqrt{n\sigma_j^2/\mu_j^3}} \le x\right) = P(N_j(n) < r),$$

where r is the smallest integer greater than $\frac{n}{\mu_j} + x\sigma_j\sqrt{n/\mu_j^3}$, i.e.,

$$r = \frac{n}{\mu_j} + x\sigma_j\sqrt{\frac{n}{\mu_j^3}} + \theta, \quad 0 < \theta \le 1.$$

Further,

$$P(N_j(n) < r) = P(\tau_j(r) > n) = P\left(\frac{\tau_j(r) - r\mu_j}{\sigma_j\sqrt{r}} > \frac{n - r\mu_j}{\sigma_j\sqrt{r}}\right)$$

and

$$\lim_{n\to\infty} \frac{n - r\mu_j}{\sigma_j\sqrt{r}} = -x.$$

Since $\tau_j(r)$ is the sum of r independent and identically distributed random variables, and $n \to \infty \iff r \to \infty$, the central limit theorem (Štěpán, 1987, Theorem IV.3.4), gives us

$$\lim_{r\to\infty} P\left(\frac{\tau_j(r) - r\mu_j}{\sigma_j\sqrt{r}} > \frac{n - r\mu_j}{\sigma_j\sqrt{r}}\right) = 1 - \Phi(-x) = \Phi(x).$$

□

Example 2.20. Consider the Markov chain $\{X_n, n \in \mathbb{N}_0\}$ with states 0 and 1, and the probability transition matrix

$$\boldsymbol{P} = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix},$$

where $0 < a < 1, \quad 0 < b < 1$. The chain is irreducible, the states are positive recurrent with the unique stationary distribution

$$\pi_0 = \frac{b}{a+b}, \quad \pi_1 = \frac{a}{a+b}.$$

The distribution of the random variable $\tau_0(1) = T_1$ and random variable $T_2 = \tau_0(2) - \tau_0(1), \ldots$, is given by Theorem 2.6. By the straightforward calculation we get

$$P(T_1 = n | X_0 = 0) = P_0(\tau_0(1) = n) = f_{00}^{(n)},$$

where

$$f_{00}^{(n)} = \begin{cases} 1-a, & n = 1, \\ ab(1-b)^{n-2}, & n \ge 2 \end{cases}$$

(notice that in this case $f_{00}^{(n)} = P(X_1 = 1, \ldots, X_{n-1} = 1, X_n = 0 | X_0 = 0)$) and similarly,

$$P(T_1 = n|X_0 = 1) = P_1(\tau_0(1) = n) = f_{10}^{(n)},$$

where

$$f_{10}^{(n)} = b(1-b)^{n-1}, \quad n \geq 1.$$

If the initial state is 0, $T_1, T_2, \ldots$ are independent and identically distributed with the distribution $\{f_{00}^{(n)}\}$. The mean and variance are

$$\mu_0 = \frac{1}{\pi_0} = \frac{a+b}{b}, \quad \sigma_0^2 = \frac{a(2-a-b)}{b^2}.$$

If the initial state is 1, T_1 has the distribution $\{f_{10}^{(n)}\}$ with the mean $\frac{1}{b}$ and the variance $\frac{1-b}{b^2}$.

The random variable $N_0(n)$ has approximately the normal distribution

$$\mathcal{N}\left(\frac{nb}{a+b}, \frac{nab(2-a-b)}{(a+b)^3}\right).$$

For $a+b=1$, this distribution simplifies to

$$N_0(n) \sim \mathcal{N}(nb, \ nb(1-b)).$$

2.8 Markov reward chains

Consider an irreducible Markov chain with a finite state space. Recall that in this case all the states are positive recurrent; suppose further that they are aperiodic (thus, all states are ergodic). Let $\boldsymbol{P} = \{p_{ij}, i, j \in S\}$ be the transition probability matrix. To any one-step transition from state i to state j associate a reward z_{ij}. Let $\boldsymbol{Z} = \{z_{ij}, i, j \in S\}$ be the matrix of rewards. Then the expected reward (earning or cost) associated with one transition from state i is $q_i = \sum_{j \in S} z_{ij} p_{ij}$.

Generally, let $v_i(n)$ denote the expected reward in n transitions (steps) given that at origin the chain was at state i, let $\boldsymbol{v}(n) = \{v_i(n), \ i \in S\}$. Put $\boldsymbol{v}(0) = \boldsymbol{0}$ and denote $\boldsymbol{q} = \{q_i, i \in S\}$.

Theorem 2.29. *The mean reward in n steps satisfies the recursive relation*

$$\boldsymbol{v}(n) = \boldsymbol{q} + \boldsymbol{P}\boldsymbol{v}(n-1), \quad n \geq 1 \ , \tag{2.44}$$

where $\boldsymbol{q} = \boldsymbol{v}(1)$ is the mean reward in one step.

Proof. Observations $(X_0 = i, X_1 = i_1, \ldots, X_n = i_n)$ yield reward $z_{ii_1} + z_{i_1i_2} + \cdots + z_{i_{n-1}i_n}$; with probability $p_i p_{ii_1} \ldots p_{i_{n-1}i_n}$, thus

$$\begin{aligned} v_i(n) &= \sum_{i_1} \cdots \sum_{i_n} (z_{ii_1} + z_{i_1i_2} + \cdots + z_{i_{n-1}i_n}) p_{ii_1} \ldots p_{i_{n-1}i_n} \\ &= \sum_{i_1} p_{ii_1} \Bigg[z_{ii_1} \sum_{i_2} \cdots \sum_{i_n} p_{i_1i_2} \ldots p_{i_{n-1}i_n} + \\ &\quad + \sum_{i_2} \cdots \sum_{i_n} (z_{i_1i_2} + \cdots + z_{i_{n-1}i_n}) p_{i_1i_2} \ldots p_{i_{n-1}i_n} \Bigg] \\ &= \sum_{i_1} p_{ii_1} z_{ii_1} + \sum_{i_1} p_{ii_1} v_{i_1}(n-1) = q_i + \sum_{i_1} p_{ii_1} v_{i_1}(n-1), \end{aligned}$$

where $q_i = \sum_j p_{ij} z_{ij} = v_i(1)$.

□

Denote

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\pi}^T \\ \vdots \\ \boldsymbol{\pi}^T \end{pmatrix}, \tag{2.45}$$

where $\boldsymbol{\pi} = \{\pi_j, j \in S\}$ is the stationary distribution.

Theorem 2.30. *Under the above assumption, the mean reward in n steps is*

$$\boldsymbol{v}(n) = (\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1}\boldsymbol{q} - (\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1}(\boldsymbol{P} - \boldsymbol{\Pi})^n\boldsymbol{q} + (n-1)\boldsymbol{\Pi}\boldsymbol{q}. \tag{2.46}$$

Proof. Using (2.44) recursively we get

$$\boldsymbol{v}(n) = \boldsymbol{q} + \boldsymbol{P}\boldsymbol{v}(n-1) = \boldsymbol{q} + \boldsymbol{P}(\boldsymbol{q} + \boldsymbol{P}\boldsymbol{v}(n-2)) = \cdots = \sum_{k=0}^{n-1} \boldsymbol{P}^k \boldsymbol{q}. \tag{2.47}$$

Since all the states are aperiodic, it holds

$$\boldsymbol{P}^k \to \boldsymbol{\Pi} \text{ pro } k \to \infty.$$

From the properties of the stationary distribution it follows that $\boldsymbol{P}\boldsymbol{\Pi} = \boldsymbol{\Pi}$, $\boldsymbol{\Pi}\boldsymbol{P} = \boldsymbol{\Pi}$ a také $\boldsymbol{\Pi}^2 = \boldsymbol{\Pi}$. Hence, we have

$$(\boldsymbol{P} - \boldsymbol{\Pi})^2 = \boldsymbol{P}^2 - \boldsymbol{\Pi}\boldsymbol{P} - \boldsymbol{P}\boldsymbol{\Pi} + \boldsymbol{\Pi}^2 = \boldsymbol{P}^2 - \boldsymbol{\Pi}$$

and by the induction for $k \geq 1$ we obtain

$$(\boldsymbol{P} - \boldsymbol{\Pi})^k = \boldsymbol{P}^k - \boldsymbol{\Pi}.$$

Since the matrix $\boldsymbol{P} - \boldsymbol{\Pi}$ is squared and satisfies

$$(\boldsymbol{P} - \boldsymbol{\Pi})^k = \boldsymbol{P}^k - \boldsymbol{\Pi} \to \mathbf{0} \text{ for } k \to \infty,$$

according to Theorem B.2 in Appendix B there is the matrix $(\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1}$ such that

$$(\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1} = \sum_{k=0}^{\infty} (\boldsymbol{P} - \boldsymbol{\Pi})^k = \boldsymbol{I} + \sum_{k=1}^{\infty} (\boldsymbol{P}^k - \boldsymbol{\Pi}).$$

Thus, the series

$$\sum_{k=0}^{\infty} (\boldsymbol{P}^k - \boldsymbol{\Pi}) = \boldsymbol{I} - \boldsymbol{\Pi} + \sum_{k=1}^{\infty} (\boldsymbol{P}^k - \boldsymbol{\Pi}) = (\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1} - \boldsymbol{\Pi}$$

is convergent and relation 2.47 can be further written as

$$\begin{aligned} \boldsymbol{v}(n) &= \sum_{k=0}^{n-1} \boldsymbol{P}^k \boldsymbol{q} = \sum_{k=0}^{n-1} (\boldsymbol{P}^k - \boldsymbol{\Pi}) \boldsymbol{q} + n \boldsymbol{\Pi} \boldsymbol{q} \qquad (2.48) \\ &= \sum_{k=0}^{\infty} (\boldsymbol{P}^k - \boldsymbol{\Pi}) \boldsymbol{q} - \sum_{k=n}^{\infty} (\boldsymbol{P}^k - \boldsymbol{\Pi}) \boldsymbol{q} + n \boldsymbol{\Pi} \boldsymbol{q} \\ &= (\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1} \boldsymbol{q} - \boldsymbol{\Pi} \boldsymbol{q} - \sum_{k=n}^{\infty} (\boldsymbol{P}^k - \boldsymbol{\Pi}) \boldsymbol{q} + n \boldsymbol{\Pi} \boldsymbol{q} \\ &= (\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1} \boldsymbol{q} - \boldsymbol{\Pi} \boldsymbol{q} - (\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1} (\boldsymbol{P} - \boldsymbol{\Pi})^n \boldsymbol{q} + n \boldsymbol{\Pi} \boldsymbol{q}. \end{aligned}$$

□

Remark. For sufficiently large n the term with $(\boldsymbol{P} - \boldsymbol{\Pi})^n$ is negligible and (2.48) can be written as

$$\boldsymbol{v}(n) \simeq (n-1) \boldsymbol{\Pi} \boldsymbol{q} + (\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1} \boldsymbol{q}.$$

Example 2.21. A new product at the market is evaluated every month either to be successful (state 0) or poor (state 1). Suppose that the sale can be modeled by a Markov chain with two states 0,1 and the transition probability matrix

$$\boldsymbol{P} = \begin{pmatrix} 0,8 & 0,2 \\ 0,3 & 0,7 \end{pmatrix}$$

and that transitions between states are associated with rewards

$$\boldsymbol{Z} = \begin{pmatrix} 10 & 5 \\ 10 & -20 \end{pmatrix}.$$

We can easily calculate

$$\boldsymbol{\Pi} = \begin{pmatrix} 0,6 & 0,4 \\ 0,6 & 0,4 \end{pmatrix}, \quad (\boldsymbol{I} - (\boldsymbol{P} - \boldsymbol{\Pi}))^{-1} = \begin{pmatrix} 1,4 & -0,4 \\ -0,6 & 1,6 \end{pmatrix}$$

and for the vector $\boldsymbol{q}$ with components $q_i = \sum_j p_{ij} z_{ij}$ we have $\boldsymbol{q} = (9, -11)^T$. The expected earning from the sale in n months is (for large n)

$$\boldsymbol{v}(n) \simeq (n + 16, n - 24)^T.$$

Discounting: Suppose now that the transition from state i at time k to state j is realized at time $k+1$ and at time k the transition is evaluated by reward z_{ij}, and discounting is considered. Then the discounted reward of the transition from i to j is $\beta^k z_{ij}$, where $0 < \beta < 1$ is the deductible (discounting factor). The discounted mean reward for n periods given the initial state i is

$$\begin{aligned} v_i(n) &= \sum_{i_1} \sum_{i_2} \cdots \sum_{i_n} (z_{ii_1} + \beta z_{i_1 i_2} + \cdots + \beta^{n-1} z_{i_{n-1} i_n}) p_{ii_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n} \\ &= \sum_{i_1} p_{ii_1} [z_{ii_1} + \beta \sum_{i_2} \cdots \sum_{i_n} (z_{i_1 i_2} + \beta z_{i_2 i_3} + \cdots + \beta^{n-2} z_{i_{n-1} i_n}) p_{i_1 i_2} \ldots p_{i_{n-1} i_n}] \\ &= \sum_{i_1} p_{ii_1} z_{ii_1} + \beta \sum_{i_1} p_{ii_1} v_{i_1}(n-1) = q_i + \beta \sum_{i_1} p_{ii_1} v_{i_1}(n-1), \end{aligned}$$

in the matrix form

$$\boldsymbol{v}(n) = \boldsymbol{q} + \beta \boldsymbol{P} \boldsymbol{v}(n-1), \quad n \geq 1. \tag{2.49}$$

Theorem 2.31. *For the discounted reward it holds*

$$\boldsymbol{v}(n) = (\boldsymbol{I} - \beta \boldsymbol{P})^{-1} \boldsymbol{q} - \beta^n (\boldsymbol{I} - \beta \boldsymbol{P})^{-1} \boldsymbol{P}^n \boldsymbol{q}$$

and

$$\lim_{n \to \infty} \boldsymbol{v}(n) = (\boldsymbol{I} - \beta \boldsymbol{P})^{-1} \boldsymbol{q}.$$

Proof. From the recursive relation (2.49) we get

$$\boldsymbol{v}(n) = \boldsymbol{q} + \beta \boldsymbol{P}\boldsymbol{v}(n-1) = \boldsymbol{q} + \beta \boldsymbol{P}(\boldsymbol{q} + \beta \boldsymbol{P}\boldsymbol{v}(n-2)) = \cdots = \sum_{k=0}^{n-1} \beta^k \boldsymbol{P}^k \boldsymbol{q}.$$

Since $0 < \beta < 1$ and $\boldsymbol{P}^k \to \boldsymbol{\Pi}$, the matrix $\beta^k \boldsymbol{P}^k$ converges to the null matrix. According to Theorems B.3 and B.2

$$\boldsymbol{v}(n) = \sum_{k=0}^{n-1} \beta^k \boldsymbol{P}^k \boldsymbol{q} = \sum_{k=0}^{\infty} \beta^k \boldsymbol{P}^k \boldsymbol{q} - \sum_{k=n}^{\infty} \beta^k \boldsymbol{P}^k \boldsymbol{q} = (\boldsymbol{I} - \beta \boldsymbol{P})^{-1} \boldsymbol{q} - \beta^n (\boldsymbol{I} - \beta \boldsymbol{P})^{-1} \boldsymbol{P}^n \boldsymbol{q}$$

and

$$\lim_{n\to\infty} \boldsymbol{v}(n) = \lim_{n\to\infty} \sum_{k=0}^{n-1} \beta^k \boldsymbol{P}^k \boldsymbol{q} = \sum_{k=0}^{\infty} \beta^k \boldsymbol{P}^k \boldsymbol{q} = (\boldsymbol{I} - \beta \boldsymbol{P})^{-1} \boldsymbol{q}.$$

□

2.9 Controlled Markov chains

A controlled Markov chain with a (finite) state space S is defined by a system of K transition probability matrix ${}_1\boldsymbol{P}, \dots, {}_K\boldsymbol{P}$. To every matrix ${}_k\boldsymbol{P}$ there belongs a reward matrix ${}_k\boldsymbol{Z}; k \in \{1, \dots, K\}$ is the control parameter. In every step we decide which transition probability matrix ${}_k\boldsymbol{P}$ and the reward matrix ${}_k\boldsymbol{Z}$ we choose; our decision can depend on the current time and the current state.Generally, our decision could depend also on all previous states of the chain, but we will not consider this general situation.

Let $k_i(n) \in \{1, \dots, K\}$ be the decision (i.e., the value of the control parameter) at time n, under the assumption that at time n the chain is in state i; in the next step we choose matrices ${}_{k_i(n)}\boldsymbol{P}, {}_{k_i(n)}\boldsymbol{Z}$. If X_n is the state of the chain at time n, then $P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = {}_{k_i(n)}p_{ij}$, and in general, the sequence $\{X_n, n \in \mathbb{N}_0\}$ is a non homogeneous Markov chain. The vector $\boldsymbol{k}(n) = \{k_i(n), i \in S\}$ is the vector of possible decisions at time n; the sequence $\{\boldsymbol{k}(n), n = 0, 1, \dots\}$ is called Markovian control of the chain. If $\boldsymbol{k}(n)$ is independent of n, we speak on the homogeneous control.

Denote by $v_i(0, N)$ the mean earning from N time periods (steps), when the initial state of the chain (at time 0) was i. Similarly, let $v_i(m, N)$ denote the mean earning from $N - m$ period, if the chain was at state i in time m. Our aim is to find a control that will maximize the mean earning from N periods.

We will proceed similarly as we did proving (2.44): If the chain is in state i at time m we choose control $k_i(m)$ and the corresponding transition and rewarding matrices and get

$$v_i(m, N) = {}_{k_i(m)}q_i + \sum_{j \in S} {}_{k_i(m)}p_{ij}\, v_j(m+1, N), \tag{2.50}$$

where

$${}_{k_i(m)}q_i = \sum_{j \in S} {}_{k_i(m)}p_{ij}\; {}_{k_i(m)}z_{ij} \tag{2.51}$$

is the mean earning from one period when the control $k_i(m)$ is used.

Theorem 2.32. *Let for every $i \in S$*

$$\widehat{v_i}(N-1, N) = \max_{1 \le k \le K} {}_kq_i,$$

$$\widehat{v_i}(m-1, N) = \max_{1 \le k \le K} \left[{}_kq_i + \sum_{j \in S} {}_kp_{ij}\, \widehat{v_j}(m, N) \right], \quad m = N-1, N-2, \ldots, 1.$$

Then $\widehat{v_i}(0, N)$ is the maximum mean earning from N periods, $i \in S$.

Proof. We will show that

$$v_i(m, N) \le \widehat{v_i}(m, N) \text{ for all } m = 0, 1, \ldots, N-1.$$

We will use the mathematical induction for $m = N-1, \ldots, 0$. Obviously,

$$v_i(N-1, N) = {}_{k_i(N-1)}q_i \le \max_{1 \le k \le K} {}_kq_i = \widehat{v_i}(N-1, N).$$

Further, if $v_i(m, N) \le \widehat{v_i}(m, N)$ for some $m < N-1$, then also

$$v_i(m-1, N) \le \widehat{v_i}(m-1, N),$$

because

$$\begin{aligned} v_i(m-1, N) &\le {}_{k_i(m-1)}q_i + \sum_{j \in S} {}_{k_i(m-1)}p_{ij}\, \widehat{v_j}(m, N) \\ &\le \max_{1 \le k \le K} \left[{}_kq_i + \sum_{j \in S} {}_kp_{ij}\, \widehat{v_j}(m, N) \right] = \widehat{v_i}(m-1, N). \end{aligned}$$

Thus, $v_i(m, N) \le \widehat{v_i}(m, N)$ for all $m = 0, 1, \ldots, N-1$, $\widehat{v_i}(0, N)$ is the maximum mean earning from N periods. Let $\widehat{k_i}(m-1), i \in S$ be the value of k, such that the value of $\widehat{v_i}(m-1, N), m = 1, \ldots, N$ is achieved. Then the control $\{\widehat{\boldsymbol{k}}(n), n = 0, 1, \ldots, N-1\}$ is the optimum.

□

Remark. More on the controlled Markov chains and the connection with dynamic programming see. e.g., in Puterman (1994).

Chapter 3

Continuous time Markov chains

3.1 Basic properties

Definition 3.1. A system of integer-valued random variables $\{X_t, t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{A}, P)$ is called to be *continuous time Markov chain* with a countable state space S, if

$$P(X_t = j | X_s = i, X_{t_n} = i_n, \dots, X_{t_1} = i_1) = P(X_t = j | X_s = i) \tag{3.1}$$

for all $i, j, i_1, \dots, i_n \in S$ and for all $0 \leq t_1 < t_2 < \cdots < t_n < s < t$, such that $P(X_s = i, X_{t_n} = i_n, \dots, X_{t_1} = i_1) > 0$.

W.l.o.g., we will assume that $S = \{0, 1, \dots\}$. Relation (3.1) is again *the Markov property.*

Denote $P(X_t = j | X_s = i)$ by $p_{ij}(s, t)$; these conditional probabilities will be called *probabilities of transition* from state i at time s to state j at time t. Similarly, probabilities $p_j(t) = P(X_t = j), j \in S$, will be called *absolute probabilities at time t* and probabilities $p_j = p_j(0) = P(X_0 = j), j \in S$, are *initial probabilities.* Obviously, $p_j(t) \geq 0$ for all $j \in S$ and $\sum_{j \in S} p_j(t) = 1, t \geq 0$.

In the next, we will consider only *homogeneous* chains with continuous time, ie., such that

$$p_{ij}(s, s + t) = p_{ij}(t), \quad s \geq 0, t > 0.$$

For every $i, j \in S$ we consider the system of transition probability functions $\{p_{ij}(t),\ t > 0\}$ such that $\sum_{j \in S} p_{ij}(t) = 1$, or equivalently, the system of transition probability matrices functions $\{\boldsymbol{P}(t),\ t > 0\}$. We will always define $p_{ij}(0) = \delta_{ij}$, i.e., $\boldsymbol{P}(0) = \boldsymbol{I}$.

Similarly as in Theorem 2.1 we can prove that all the finite-dimensional distribution of a homogeneous continuous time Markov chain $\{X_t,\ t \geq 0\}$ are determined by the vector of initial probabilities $\boldsymbol{p}(0) = \{p_i(0), i \in S\}$ and the system of transition probabilities matrices functions $\{\boldsymbol{P}(t), t \geq 0\}$: for any $0 < t_1 < t_2 < \cdots < t_n$ and any $i_0, i_1, \ldots, i_n \in S$ it holds

$$P(X_0 = i_0, X_{t_1} = i_1, \ldots, X_{t_n} = i_n) = p_{i_0}(0) p_{i_0 i_1}(t_1) p_{i_1 i_2}(t_2 - t_1) \ldots p_{i_{n-1} i_n}(t_{i_n} - t_{i_{n-1}}), \tag{3.2}$$

from which we immediately get

$$p_j(t) = P(X_t = j) = \sum_{i \in S} p_i(0) p_{ij}(t), \quad j \in S, \tag{3.3}$$

or, in the matrix form,

$$\boldsymbol{p}(t)^T = \boldsymbol{p}(0)^T \boldsymbol{P}(t).$$

Analogously to discrete time Markov chains it can be proved that for every $s \geq 0, t \geq 0$

$$p_{ij}(s+t) = \sum_{k \in S} p_{ik}(s) p_{kj}(t), \quad i, j \in S, \tag{3.4}$$

i.e., in the matrix form,

$$\boldsymbol{P}(s+t) = \boldsymbol{P}(s)\boldsymbol{P}(t),$$

which is again the *Chapman-Kolmogorov equation.*

On the other hand, it can be also shown that to any probability vector $\boldsymbol{p} = \{p_i \geq 0, \sum_{i \in S} p_i = 1\}$ and to any system of stochastic matrix functions $\{\boldsymbol{P}(t), t \geq 0\}$ that satisfy the Chapman-Kolmogorov equation, there exists a homogeneous time continuous Markov chain the initial probabilities of which are given by the vector $\boldsymbol{p}$ and the system of transition probability matrix functions is just $\{\boldsymbol{P}(t),\ t \geq 0\}$ (see, e.g., Chung, 1967, pp. 141–142).

In our next considerations we will always assume that

$$\lim_{t \to 0_+} p_{ij}(t) = \delta_{ij} \quad i, j \in S, \tag{3.5}$$

which together with the assumption $p_{ij}(0) = \delta_{ij}$ means that transition probabilities $p_{ij}(t)$ are right-continuous at 0. Moreover we can prove the following

Theorem 3.1. *Under assumption (3.5), the transition probabilities $p_{ij}(t),\ t \geq 0$ are uniformly continuous.*

Proof. According to the Chapman-Kolmogorov equation (3.4), for every $h > 0$

$$p_{ij}(t+h) - p_{ij}(t) = \sum_{k \neq i} p_{ik}(h) p_{kj}(t) - p_{ij}(t)(1 - p_{ii}(h)),$$

further

$$\sum_{k \neq i} p_{ik}(h) p_{kj}(t) \leq \sum_{k \neq i} p_{ik}(h) = 1 - p_{ii}(h),$$

so that

$$|p_{ij}(t+h) - p_{ij}(t)| \leq |1 - p_{ii}(h)| + |p_{ij}(t)(1 - p_{ii}(h))| \leq 2(1 - p_{ii}(h)). \tag{3.6}$$

From (3.5) and (3.6) it follows that $|p_{ij}(t+h) - p_{ij}(t)|$ converges to zero as $h \to 0_+$ independently of t. Similarly we proceed with $|p_{ij}(t) - p_{ij}(t-h)|$. □

Remark. Under the same assumptions and as a consequence of Theorem 3.1 we can prove that the absolute probabilities $p_j(t)$ are uniformly continuous for $t \geq 0$.

Theorem 3.2. *A homogeneous time continuous Markov chain $\{X_t, t \geq 0\}$ that satisfies (3.5), is stochastically continuous.*

Proof. Since $P(|X_t - X_s| > \epsilon) \leq P(X_t \neq X_s) = 1 - P(X_t = X_s)$ for every t and s, it suffices to show that for every $t \geq 0$

$$\lim_{h \to 0_+} P(X_{t+h} = X_t) = 1, \quad \lim_{h \to 0_+} P(X_{t-h} = X_t) = 1.$$

Obviously,

$$P(X_{t+h} = X_t) = \sum_{j \in S} P(X_t = j, X_{t+h} = j) = \sum_{j \in S} p_j(t) p_{jj}(h),$$

and since $|p_j(t) p_{jj}(h)| \leq p_j(t)$ and $\sum_{j \in S} p_j(t) = 1$, we have

$$\lim_{h \to 0_+} P(X_{t+h} = X_t) = \sum_{j \in S} p_j(t) \lim_{h \to 0_+} p_{jj}(h) = 1.$$

In the rest of the proof we proceed analogously.

□

If $\{X_t\}$ is stochastically continuous then, according to Theorem 1.2, there exists its version that is separable and measurable. In the sequel, we will always assume that $\{X_t\}$ is separable and measurable.

Further, it can be proved that trajectories of a separable continuous time Markov chain are almost surely step functions (Doob, 1953, Chap. VI, \$ 1). We will always assume that they are right-continuous.

Now, let us consider transition probability functions.

Theorem 3.3. *For every $i \in S$ there exists the limit*

$$\lim_{h \to 0_+} \frac{1 - p_{ii}(h)}{h} := q_i \leq \infty, \tag{3.7}$$

for every $i, j \in S,\ i \neq j$ there exist the limits

$$\lim_{h \to 0_+} \frac{p_{ij}(h)}{h} := q_{ij} < \infty, \tag{3.8}$$

and for every $i \in S$

$$\sum_{j \neq i} q_{ij} \leq q_i.$$

Proof. The Proof of assertions (3.7), (3.8) is given in Chung (1967), Theorem II.2.4 and Theorem II.2.5, or in Karlin and Taylor (1981), Chap. 14, Theorems 1.1, 1.2. Utilizing the properties of stochastic matrices we have for every $h \geq 0$ and $n \in \mathbb{N}$

$$\frac{1 - p_{ii}(h)}{h} = \sum_{\substack{j \in S \\ j \neq i}} \frac{p_{ij}(h)}{h} \geq \sum_{\substack{j=0 \\ j \neq i}}^{N} \frac{p_{ij}(h)}{h}$$

and using the limit as $h \to 0_+$ and then as $N \to \infty$, we get the last assertion.

□

In the next we will consider only such processes that satisfy

$$q_i = \sum_{j \neq i} q_{ij} \text{ for all } i \in S. \tag{3.9}$$

Remark. For a process with finite state space S, (3.9) always holds since

$$0 = 1 - \sum_{j \in S} p_{ij}(h), \quad h \geq 0,$$

and

$$0 = \lim_{h \to 0_+} \frac{1 - \sum_j p_{ij}(h)}{h} = \lim_{h \to 0_+} \frac{1 - p_{ii}(h) - \sum_{j \neq i} p_{ij}(h)}{h} = q_i - \sum_{j \neq i} q_{ij}.$$

Definition 3.2. The non-negative numbers q_{ij} defined in Theorem (3.3) are called *the transition intensities* from state i to state j, the non-negative number q_i is called *the total intensity* . The matrix $\boldsymbol{Q} = \{q_{ij}, i, j \in S\}$, where $q_{ii} = -q_i$, is called *the transition intensity matrix.*

Example 3.1. *Poisson process.* Let us recall the process from Example 1.4. We consider events of the same type that occur randomly in time. The numbers of events that occur in disjoint time intervals are supposed to be independent random variables the distribution of which depends only on the length of these intervals. Suppose that exactly one event occurs within the interval $(t, t+h]$ with probability $\lambda h + o(h)$ for some $\lambda > 0$ and all t, more that one event occur in $(t, t+h]$ with probability $o(h)$ and no event with probability $1 - \lambda h + 0(h)$; the symbol $o(h)$ means that $o(h)/h \to 0$ as $h \to 0_+$. Put $X_0 = 0$ and let $X_t, t > 0$ denote the number of events that occur within the interval $(0, t]$. Then $\{X_t, t \geq 0\}$ is a stochastic process of integer-valued random variables. From our assumptions it holds that for arbitrary time instants $t_1 < t_2 < t_3 < t_4$ the numbers of events $X_{t_2} - X_{t_1}$ and $X_{t_4} - X_{t_3}$ that occur in intervals $(t_1, t_2]$ a $(t_3, t_4]$ are independent random variables. We say that $\{X_t, t \geq 0\}$ has *independent increments.* From here the Markov property of $\{X_t, t \geq 0\}$ also follows. The distribution of the increment $X_t - X_s$ depends only on the length of interval $(s, t]$; we say that $\{X_t, t \geq 0\}$ has *stationary increments.* Further, for any $t \geq 0$

$$\begin{aligned} P(X_{t+h} = j | X_t = i) = p_{ij}(h) &= \lambda h + o(h), & j &= i+1, \\ &= 1 - \lambda h + o(h), & j &= i, \\ &= o(h), & j &> i+1, \\ &= 0, & j &< i. \end{aligned}$$

The transition intensities are $q_{i,i+1} = \lambda, q_i = -q_{ii} = \lambda, q_{ij} = 0$ otherwise; the matrix $\boldsymbol{Q}$ is

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -\lambda & \lambda & 0 & \dots \\ & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

The above described process is called the *Poisson process with intensity* λ. Later on we will show that the absolute probabilities of this process are probabilities of the Poisson distribution with the parameter λ.

Example 3.2. *Linear birth and death process.* Consider a population, the individuals of which can reproduce and die. During an interval $(t, t+h]$ each individual can give a birth of a new one with probability $\lambda h + o(h)$, for $\lambda > 0$, more than one with probability $o(h)$. Each individual can die during interval $(t, t+h]$ with probability $\mu h + o(h)$, $\mu > 0$. All individuals behave independently. If the population size at time t is X_t, then for $h \to 0+$

$$\begin{aligned} P(X_{t+h} = j+1 | X_t = j) &= j\lambda h + o(h), \\ P(X_{t+h} = j-1 | X_t = j) &= j\mu h + o(h), \\ P(X_{t+h} = k | X_t = j) &= o(h), \quad |k| > 1, \\ P(X_{t+h} = j | X_t = j) &= 1 - j\lambda h - j\mu h + o(h) \end{aligned}$$

and $\{X_t, t \geq 0\}$ is a homogeneous continuous time Markov chain with the state space $S = \{0, 1, \ldots, \}$, with intensities

$$\begin{aligned} q_{j,j+1} &= j\lambda, \quad j = 0, 1, \ldots, \\ q_{j,j-1} &= j\mu, \quad j = 1, \ldots, \\ q_{ij} &= 0, \qquad i \neq j \text{ otherwise }, \\ q_j &= j(\lambda + \mu). \end{aligned}$$

Thus, the intensity matrix $\boldsymbol{Q}$ is

$$\boldsymbol{Q} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \ldots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \ldots \\ 0 & 2\mu & -2(\lambda+\mu) & 2\lambda & 0 & \ldots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \end{pmatrix}.$$

Theorem 3.4. *Let $\{X_t, t \geq 0\}$ be a homogeneous time continuous Markov chain with a countable state space. Then for all $s \geq 0$ and all $h > 0$*

$$P(X_t = i, s \leq t \leq s+h | X_s = i) = e^{-q_i h} \tag{3.10}$$

(where $e^{-q_i h} = 0$, if $q_i = \infty$).

Proof. Since $\{X_t\}$ is stochastically continuous and separable, it holds (Doob, 1953, p. 245)

$$P(X_t = i, s \leq t \leq s+h) = \lim_{n\to\infty} P(X_{t_{0,n}} = i, X_{t_{1,n}} = i, \ldots, X_{t_{n,n}} = i) \tag{3.11}$$

for every sequence $D_n = \{t_{0,n} < \cdots < t_{n,n}\}$ of divisions of the interval $[s, s+h]$ the norm of which converges to zero as $n \to \infty$.

Put $D_n = \{t_{k,n} : t_{k,n} = s + \frac{hk}{2^n},\ k = 0, 1, \dots, 2^n\},\ n = 1, 2 \dots$. From (3.11) and (3.2) we have

$$P(X_t = i, s \le t \le s+h | X_s = i) = \frac{P(X_t = i, s \le t \le s+h)}{P(X_s = i)}$$
$$= \lim_{n\to\infty} \frac{P(X_{t_{0,n}} = i, X_{t_{1,n}} = i, \dots, X_{t_{n,n}} = i)}{P(X_s = i)} = \lim_{n\to\infty} \left[p_{ii}\left(\frac{h}{2^n}\right)\right]^{2^n}.$$

If $q_i < \infty$, then due to (3.7) we can write

$$\lim_{n\to\infty} \left[p_{ii}\left(\frac{h}{2^n}\right)\right]^{2^n} = \lim_{n\to\infty} \left[1 - q_i \frac{h}{2^n} + o\left(\frac{h}{2^n}\right)\right]^{2^n} = e^{-q_i h}.$$

If $q_i = \infty$, then (3.7) implies that for any $A > 0$, $p_{ii}(\delta) \le 1 - A\delta \le e^{-A\delta}$, for a sufficiently small δ; from here we have

$$\lim_{n\to\infty} \left[p_{ii}\left(\frac{h}{2^n}\right)\right]^{2^n} \le e^{-Ah}.$$

Since A can be chosen to be arbitrarily large, we get the result.

□

Theorem 3.5. *If $q_i = 0$, then $p_{ii}(t) = 1$ for all $t \ge 0$. If $0 < q_i < \infty$, then time the chain stays at state i (holding time at i) is a random variable with the exponential distribution and the mean value $\frac{1}{q_i}$.*

Proof. Obviously,

$$p_{ii}(t) = P(X_t = i | X_0 = i) \ge P(X_s = i, 0 \le s \le t | X_0 = i) = e^{-q_i t}$$

for every $t > 0$. Thus, for $q_i = 0$, $p_{ii}(t) = 1$ for all $t \ge 0$.

Now, let $0 < q_i < \infty$ and let $\tau = \inf\{t > 0 : X_t \ne i\}$ be the first exit time from state i. Then, the holding time T at state i satisfies

$$P_i(T > s) = P_i(\tau > s) = P(X_t = i, 0 \le t \le s | X_0 = i) = e^{-q_i s},$$

thus, T has the exponential ditribution with parameter q_i and density

$$f(x) = \begin{cases} q_i e^{-q_i x}, & x \ge 0 \\ 0 & \text{otherwise;} \end{cases}$$

therefore, the mean holding time at state i is $\frac{1}{q_i}$.

□

Definition 3.3. State $i \in S$ such that $q_i = 0$ is called *absorbing.* State $i \in S$ such that $0 < q_i < \infty$ is called *stable.* State i is called *instantaneous* if $q_i = \infty$.

If the chain enters an absorbing state, it stays at this state forever; the mean holding time at the absorbing state is infinitely large. The mean holding time in an instantaneous state is zero.

Definition 3.4. Let $\{X_t, t \geq 0\}$ be a stochastic process defined on $(\Omega, \mathcal{A}, P)$, with a countable state space S having right-continuous trajectories. Let $\mathcal{F}_t$ be the σ-field generated by the family $\{X_s, s \leq t\}$, i.e., $\mathcal{F}_t = \sigma\{X_s, s \leq t\}$. Random variable $\tau : \Omega \to [0, \infty]$ is called *the stopping time* of the process $\{X_t, t \geq 0\}$, if $[\tau \leq t] \in \mathcal{F}_t$ for every $t \geq 0$.

The stopping time is a random variable that is independent of future values of the stochastic process. For example, the first exit time from state j of a continuous time Markov chain, $\tau_j = \inf\{t \geq 0, X_t \neq j\}$, is the stopping time. Due to separability, we have for a countable set Q dense in $[0, \infty)$)

$$[\tau_j > t] = [X_s = j, 0 \leq s \leq t] = \bigcap_{s \in Q \cap [0,t]} [X_s = j] \in \mathcal{F}_t,$$

thus $[\tau_j \leq t] \in \mathcal{F}_t$. Let

$$\mathcal{F}_\infty = \sigma\left(\bigcup_{t \geq 0} \mathcal{F}_t\right) = \sigma\{X_t, t \geq 0\}.$$

Obviously, $\mathcal{F}_t \subset \mathcal{F}_\infty \subset \mathcal{A}$.

Theorem 3.6. *Let τ be a stopping time of the Markov chain $\{X_t, t \geq 0\}$. Then X_τ is $\mathcal{F}_\tau-$measurable random variable where*

$$\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap [\tau \leq t] \in \mathcal{F}_t, t \geq 0\}.$$

Proof. It suffices to show that for every $i \in S$ and every $t \geq 0$

$$[X_\tau = i] \cap [\tau \leq t] \in \mathcal{F}_t.$$

Obviously, it holds $[X_\tau = i] \cap [\tau \leq t] = ([X_t = i] \cap [\tau = t]) \cup ([X_\tau = i] \cap [\tau < t])$. Since $\{X_t\}$ takes only non-negative integers and has right-continuous trajectories we have

$$\begin{aligned}
[X_\tau = i] \cap [\tau \leq t] &= [X_t = i] \cap [\tau = t] \\
&\cup \left(\bigcup_{n=0}^{\infty} \bigcap_{m=n}^{\infty} \bigcup_{k=1}^{z[2^m t]} [X_{k2^{-m}} = i] \cap [(k-1)2^{-m} \leq \tau < k2^{-m}]\right) \in \mathcal{F}_t,
\end{aligned}$$

where $z[x]$ denotes the integer part of x. □

Theorem 3.7. *Let $0 < q_i < \infty$, and let τ_i be the first exit time from state i. Then*

$$P(X_{\tau_i} = j | X_0 = i) = \frac{q_{ij}}{q_i}$$

for all $j \neq i$, i.e., the probability that the chain hits state j just after leaving the initial state i is equal to $\frac{q_{ij}}{q_i}$.

Proof. Denote

$$A \quad = \quad [X_0 = i,\ X_{\tau_i} = j,\ 0 < \tau_i < \infty]$$

and consider sets

$$\begin{aligned} A_n &= \bigcup_{k=1}^{\infty} A_{n,k} \quad \text{pro } n \in \mathbb{N}, \text{ where} \\ A_{n,k} &= \{\omega : X_t(\omega) = i,\ 0 \leq t \leq k2^{-n},\ X_{(k+1)2^{-n}}(\omega) = j\} \quad \text{pro } k \in \mathbb{N}. \end{aligned}$$

First, we will show that $A = \lim_{n\to\infty} A_n$.

1. Let $\omega \in A$.

   Then $0 < \tau_i(\omega) < \infty$, $X_{\tau_i(\omega)}(\omega) = j$ and $X_t(\omega) = i$ for all $0 \leq t < \tau_i(\omega)$.

   Since the trajectories of the process $\{X_t, t \geq 0\}$ are right-continuous, there is $\delta > 0$ such that $X_s(\omega) = j$ for all $\tau_i(\omega) \leq s < \tau_i(\omega) + \delta$. Then to any $n \in \mathbb{N}$ that satisfies $2^{-n} \leq \delta$ there exists the unique $k_n \in \mathbb{N}$ such that

   $$k_n 2^{-n} < \tau_i(\omega) \leq (k_n + 1)2^{-n} < \tau_i(\omega) + \delta$$

   and

   $$X_t(\omega) = i,\ 0 \leq t \leq k_n 2^{-n}, \quad X_{(k_n+1)2^{-n}}(\omega) = j.$$

   Thus, $\omega \in A_{n,k_n}$, and $\omega \in A_n$ for all $n \in \mathbb{N}$ that fulfill $2^{-n} \leq \delta$. From here,

   $$\liminf_{n\to\infty} A_n \supset A.$$

2. Let $\omega \in \limsup_{n\to\infty} A_n$.

   Then there exists a sequence of integers $1 < n_1 < n_2 < n_3 < \ldots.$ and associated $\kappa_1, \kappa_2, \kappa_3, \ldots. \in \mathbb{N}$ such that for every $k \in \mathbb{N}$,

   $$X_t(\omega) = i \quad \forall\, 0 \leq t \leq \kappa_k 2^{-n_k}, \quad X_{(\kappa_k+1)2^{-n_k}}(\omega) = j.$$

Then the following inequalities hold

$$\kappa_k 2^{-n_k} \le \kappa_{k+1} 2^{-n_{k+1}} < \tau_i(\omega) \le (\kappa_{k+1}+1)2^{-n_{k+1}} \le (\kappa_k+1)2^{-n_k}.$$

Therefore we have

$$\tau_i(\omega) = \sup_{k\in\mathbb{N}} \kappa_k 2^{-n_k} = \inf_{k\in\mathbb{N}} (\kappa_k+1)2^{-n_k}.$$

From here

$$\tau_i(\omega) > 0\,, \quad X_t(\omega) = i \quad \forall\, 0 \le t < \tau_i(\omega)$$

and due to the right-continuity of the trajectories,

$$X_{\tau_i(\omega)}(\omega) = j,$$

thus $\omega \in A$.

It means that

$$\limsup_{n\to\infty} A_n \subset A.$$

We have realized that $\lim_{n\to\infty} A_n = A$, therefore the set A is measurable. For given $n \in \mathbb{N}$ the sets $A_{n,k}$, $k \in \mathbb{N}$ are disjoint. From the continuity of probability measure and by 3.10 it follows

$$\begin{aligned} P(A|X_0=i) &= \lim_{n\to\infty} P(A_n|X_0=i) = \lim_{n\to\infty} \sum_{k=1}^{\infty} P(A_{n,k}|X_0=i) \\ &= \lim_{n\to\infty} \sum_{k=1}^{\infty} \exp\{-kq_i 2^{-n}\} p_{ij}(2^{-n}) \\ &= \lim_{n\to\infty} p_{ij}(2^{-n})[1-\exp\{-q_i 2^{-n}\}]^{-1} \exp\{-q_i 2^{-n}\} = \frac{q_{ij}}{q_i}. \end{aligned}$$

□

Due to homogeneity, the probability that the chain occupying state i at some time $t > 0$, in the interval $(t,\ \infty)$ moves first to state j is also q_{ij}/q_i.

We assume that at the origin the chain is at state i; if $q_i > 0$, it stays there for a random holding time T_1 and then moves to state j with probability $\frac{q_{ij}}{q_i}$. If $q_j > 0$, it stays at state j random holding time T_2; if $q_j = 0$, the chain stays at state j forever $(T_2 = \infty)$

etc. Let $J_0 = 0$ and $J_1 < J_2 < \ldots$ are time instants in which the transitions between the states occur, i.e.

$$\begin{aligned} J_1 &= \inf\{t > 0 : X_t \neq X_0\}, \\ J_2 &= \inf\{t > J_1 : X_t \neq X_{J_1}\}, \\ &\ldots \\ J_{n+1} &= \inf\{t > J_n : X_t \neq X_{J_n}\},\ n \geq 0. \end{aligned}$$

Holding times between transitions are $T_1 = J_1, T_2 = J_2 - J_1$ (if $J_1 < \infty$, otherwise $T_2 = \infty$) etc.; obviously,

$$J_n = \sum_{k=1}^{n} T_k, \quad n \geq 1.$$

Denote

$$\xi = \sup_{n\in\mathbb{N}} J_n = \sum_{k=1}^{\infty} T_k.$$

Definition 3.5. The homogeneous time continuous Markov chain with stable states is called *regular* if

$$P_i(\xi = \infty) = 1\ \forall\ i \in S\ .$$

The random variable ξ is called *explosion time.*

Trajectories of a chain with stable states are right-continuous, step-wise functions with the jumps at transition times $J_1, J_2, \ldots$. If the process is regular, in any finite interval only a finite number of transitions between the states occurs with probability one. If in a finite time interval infinitely many transitions occur we say that the process is explosive. If $\xi < \infty$, we define $X_t = \infty$ for $t \geq \xi$.

Now, let us define matrix $\boldsymbol{Q}^* = \{q^*_{ij}, i, j \in S\}$ by

$$q^*_{ij} = \begin{cases} \frac{q_{ij}}{q_i}, & q_i > 0 \\ 0, & q_i = 0 \end{cases} \quad i \neq j \tag{3.12}$$

$$q^*_{ii} = \begin{cases} 0, & q_i > 0 \\ 1, & q_i = 0. \end{cases}$$

Consider again the sequence of jumps $J_1, J_2, \ldots$, and define the sequence

$$\begin{aligned} Y_0 &= X_0, \\ Y_n &= X_{J_n}, \quad n = 1, 2, \ldots \end{aligned} \tag{3.13}$$

(if $J_n = \infty$, we define$Y_\infty = Y_{J_{n-1}}$).

Since J_n are stopping times, Y_n are random variables according to Theorem 3.6. Then the following assertion can be proved.

Theorem 3.8. *Let $\{X_t, t \geq 0\}$ be a regular continuous time Markov chain with the state space S and the transition intensity matrix $\boldsymbol{Q}$. Let $\{Y_n, n \in \mathbb{N}_0\}$ be the sequence of random variables as given by (3.13). Then $\{Y_n, n \in \mathbb{N}_0\}$ is a homogeneous discrete time Markov chain with the same state space S and the transition probability matrix $\boldsymbol{Q}^*$. For every $n \geq 1$ and given $Y_0, \ldots, Y_{n-1}$, the holding times $T_1, \ldots, T_n$ are independent random variables exponentially distributed with respective parameters $q_{Y_0}, \ldots, q_{Y_{n-1}}$.*

Proof. Gichman and Skorochod (1973, Vol II) or Norris (1997), Theorem 2.8.4. □

Definition 3.6. The discrete time Markov chain $\{Y_n, n \in \mathbb{N}_0\}$ is called *the embedded Markov chain of jumps* associated with the process $\{X_t, t \geq 0\}$.

The following relations hold

$$P(X_t = i) = \sum_{n=0}^{\infty} P(Y_n = i, J_n \leq t < J_{n+1})$$

and

$$P(X_t = i \text{ for some } t \geq 0) = P(Y_n = i \text{ for some } n \geq 0).$$

Example 3.3. *Explosive birth process.* Let us consider a time continuous Markov chain with the state space $S = \{1, 2, \ldots, \}$, given by the transition intensities

$$q_i = -q_{ii} = 2^i, \quad q_{i,i+1} = 2^i, \quad q_{ij} = 0 \text{ otherwise.}$$

Suppose that the initial state is 1. The holding time at this state is the random variable T_1, that is exponentially distributed with the mean value $\frac{1}{q_1} = \frac{1}{2}$, then the chain jumps to state 2 with probability one, the holding time at state 2 is independent of T_1 and exponentially distributed with the mean $\frac{1}{q_2} = \frac{1}{4}$ etc., thus the explosion time is $\xi = \sum_{k=1}^{\infty} T_k$ and

$$E\xi = \sum_{k=1}^{\infty} ET_k = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

thus

$$P_1(\xi < \infty) = 1;$$

if $P_1(\xi = \infty) > 0$ then also $E\xi = \infty$. The chain is not regular; infinitely many transitions can occur at a finite time interval with probability one.

The state space of the embedded Markov chain is $S = \{1, 2, \dots\}$ and the transition probability matrix is of the form

$$\boldsymbol{Q}^* = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 1 & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

Theorem 3.9. *A homogeneous continuous time Markov chain with stable states* $i \in S = \{0, 1, \dots\}$ *and the associated embedded Markov chain* $\{Y_n, n \in \mathbb{N}_0\}$ *is regular if and only if*

$$P_i\left(\sum_{k=0}^{\infty} \frac{1}{q_{Y_k}} = \infty\right) = 1 \qquad \forall\ i \in S, \tag{3.14}$$

where q_j *denotes the total intensity of state* j*,* $j \in S$*.*

Proof. Gichman and Skorochod (1973, Vol. II), Theorem 3 in Chap. III, Par. 2. or Resnick (1992), Proposition 5.3.1.

□

Remark. Regularity condition (3.14) is fulfilled if, e.g., $q_i \le C \quad \forall\ i \in S$, since in this case

$$\sum_{k=0}^{\infty} \frac{1}{q_{Y_k}} \ge \sum_{k=0}^{\infty} \frac{1}{C} = \infty$$

with probability 1. Obviously, if the state space is finite, the process $\{X_t, t \ge 0\}$ is regular.

Further it holds that the process, the embedded chain of which is irreducible with recurrent states, is regular. Consider some state i; since it is recurrent, the chain visits it infinitely many times at instants $n_1, n_2 \dots$, thus

$$\sum_{k=0}^{\infty} \frac{1}{q_{Y_k}} \ge \sum_{j=1}^{\infty} \frac{1}{q_{Y_{n_j}}} = \sum_{j=1}^{\infty} \frac{1}{q_i} = \infty.$$

3.2 Kolmogorov differential equations

Let us consider a continuous time Markov chain with the space state S. Then we can prove the following theorem.

Theorem 3.10. (Kolmogorov differential equations). *Suppose that $q_i < \infty$ for all $i \in S$ and (3.9) holds. Then the transition probabilities $p_{ij}(t)$ are differentiable for all $i, j \in S$ and $t > 0$ and*

$$p'_{ij}(t) = -q_i p_{ij}(t) + \sum_{k \neq i} q_{ik} p_{kj}(t) = \sum_{k \in S} q_{ik} p_{kj}(t) \tag{3.15}$$

(backward equations).

If the convergence $\frac{p_{ij}(h)}{h} \to q_{ij}$ is uniform in i for a fixed j, then for every $i, j \in S$ and $t > 0$

$$p'_{ij}(t) = -p_{ij}(t) q_j + \sum_{k \neq j} p_{ik}(t) q_{kj} = \sum_{k \in S} p_{ik}(t) q_{kj} \tag{3.16}$$

(forward equations).

In the matrices, the system of the backward equations is of the form

$$\boldsymbol{P}'(t) = \boldsymbol{Q}\boldsymbol{P}(t)$$

and the system of the forward equations reads as

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{Q}.$$

Proof. We start with the Chapman-Kolmogorov equation

$$p_{ij}(s+t) = \sum_{k=0}^{\infty} p_{ik}(s) p_{kj}(t) = \sum_{k=0}^{N} p_{ik}(s) p_{kj}(t) + \sum_{k=N+1}^{\infty} p_{ik}(s) p_{kj}(t), \quad N > i,$$

from here we have

$$\frac{p_{ij}(s+t) - p_{ij}(t)}{s} = \frac{[p_{ii}(s) - 1] p_{ij}(t)}{s} + \sum_{\substack{k=0 \\ k \neq i}}^{N} \frac{p_{ik}(s) p_{kj}(t)}{s} + h_N(s,t),$$

where

$$h_N(s,t) = \sum_{k=N+1}^{\infty} \frac{p_{ik}(s) p_{kj}(t)}{s}.$$

For all $s > 0, t \geq 0$ we get

$$0 \leq h_N(s,t) \leq \sum_{k=N+1}^{\infty} \frac{p_{ik}(s)}{s} = \frac{1 - \sum_{k=0}^{N} p_{ik}(s)}{s} = \frac{1 - p_{ii}(s)}{s} - \sum_{\substack{k=0 \\ k \neq i}}^{N} \frac{p_{ik}(s)}{s},$$

so that

$$\begin{aligned} \frac{[p_{ii}(s) - 1]p_{ij}(t)}{s} + \sum_{\substack{k=0 \\ k \neq i}}^{N} \frac{p_{ik}(s)p_{kj}(t)}{s} & \\ \leq \frac{p_{ij}(s+t) - p_{ij}(t)}{s} & \\ \leq \frac{[p_{ii}(s) - 1]p_{ij}(t)}{s} + \sum_{\substack{k=0 \\ k \neq i}}^{N} \frac{p_{ik}(s)p_{kj}(t)}{s} + \frac{1 - p_{ii}(s)}{s} - \sum_{\substack{k=0 \\ k \neq i}}^{N} \frac{p_{ik}(s)}{s}. & \end{aligned}$$

The limit for $s \to 0_+$ is

$$\begin{aligned} -q_i p_{ij}(t) + \sum_{\substack{k=0 \\ k \neq i}}^{N} q_{ik} p_{kj}(t) & \\ \leq p'_{ij}(t_+) & \\ \leq -q_i p_{ij}(t) + \sum_{\substack{k=0 \\ k \neq i}}^{N} q_{ik} p_{kj}(t) + q_i - \sum_{\substack{k=0 \\ k \neq i}}^{N} q_{ik}, & \end{aligned}$$

where $p'_{ij}(t_+)$ denotes the right-derivative of p_{ij} at t, and by using the limit $N \to \infty$ we realize that $p'_{ij}(t_+)$ satisfies (3.15). In the same way we can prove that (3.15) hold also for the left-derivative; here we use the uniform continuity of $p_{ij}(t)$ at t.

Similarly we develop the forward equations: Again, from the Chapman-Kolmogorov equation we have

$$\begin{aligned} \frac{p_{ij}(s+t) - p_{ij}(s)}{t} &= \frac{\sum_{k=0}^{\infty} p_{ik}(s)p_{kj}(t) - p_{ij}(s)}{t} \\ &= \frac{p_{ij}(s)(p_{jj}(t) - 1)}{t} + \sum_{\substack{k=0 \\ k \neq j}}^{\infty} \frac{p_{ik}(s)p_{kj}(t)}{t} \end{aligned}$$

and applying limit $t \to 0_+$ (due to the uniform convergence assumption)

$$p'_{ij}(s_+) = -p_{ij}(s)q_j + \sum_{\substack{k=0 \\ k \neq j}}^{\infty} p_{ik}(s)q_{kj} = \sum_{k=0}^{\infty} p_{ik}(s)q_{kj}.$$

Since $p_{ij}(t)$ are continuous and have continuous right-derivatives, they are differentiable (e.g., Billingsley, 1968, p. 155). □

Remark. If the state space is finite, the uniform convergence of $p_{ij}(h)/h$ in Theorem 3.10 is satisfied.

In the backward equation, the derivatives $p'_{ij}(t)$ are expressed by all the transitions to state j, while in the forward equations, all possible transition from state i are considered.

We have seen that the transition probabilities of a continuous time Markov chain satisfy systems of backward of forward differentiable equations. Now, let us consider a system of differential equations of type (3.15) or (3.16); the question arises if a solution of such a system can be a system of transition probabilities of a continuous time Markov chain. We will study only finite case.

Theorem 3.11. *Let $\boldsymbol{Q} = \{q_{ij}, 0 \leq i, j \leq N\}$ be a matrix, elements of which satisfy*

$$q_{ij} \geq 0, \;\; i \neq j, \quad q_{ii} = -\sum_{i \neq j} q_{ij}.$$

Then there exists the unique solution of equations (3.15) a (3.16), the same for both systems, that satisfies the initial condition $\boldsymbol{P}(0) = \boldsymbol{I}$, and that represents the system of transition probabilities of a continuous time Markov chain with the state space $S = \{0, 1, \ldots, N\}$. In matrices, we have solution $\boldsymbol{P}(t) = e^{\boldsymbol{Q}t}$, where $e^{\boldsymbol{Q}t}$ is the matrix exponential function defined by

$$e^{\boldsymbol{Q}t} = \sum_{k=0}^{\infty} \frac{\boldsymbol{Q}^k t^k}{k!}.$$

Proof. Systems (3.15) and (3.16) are systems of linear differential equations of the first order, with constant coefficients; the general solution of (3.15) is $\boldsymbol{P}(t) = \boldsymbol{e}^{\boldsymbol{Q}t}\boldsymbol{P}(0)$ and that of (3.16) is $\boldsymbol{P}(t) = \boldsymbol{P}(0)e^{\boldsymbol{Q}t}$. With the initial condition $\boldsymbol{P}(0) = \boldsymbol{I}$ both the systems have the solution

$$\boldsymbol{P}(t) = e^{\boldsymbol{Q}t}. \tag{3.17}$$

This solution satisfies the Chapman-Kolmogorov equation (3.4), since from the properties of the matrix exponential function we get $e^{\boldsymbol{Q}(t+s)} = e^{\boldsymbol{Q}t}e^{\boldsymbol{Q}s}$ for every $t, s \geq 0$.

Further we need to show that $\boldsymbol{P}(t)$ is a stochastic matrix for every $t \geq 0$. Summing (3.16) for all j we get

$$\sum_j p'_{ij}(t) = \sum_j \sum_k p_{ik}(t) q_{kj} = \sum_k p_{ik}(t) \sum_j q_{kj} = 0,$$

since $\sum_j q_{kj} = 0$. Hence, $\sum_j p_{ij}(t)$ is a constant for every $t \geq 0$, and since $\sum_j p_{ij}(0) = 1$, this constant equals one.

Next, we need to prove that all $p_{ij}(t)$ are non-negative. First, $p_{ii}(0) = 1$ and due to continuity of the solution, $p_{ii}(t) > 0$ at some right-neighbourhood of zero.

If $q_{ij} > 0$ for all $i \neq j$, we get $p'_{ij}(0) = q_{ij} > 0$, thus $p_{ij}(t) > 0$ for $0 < t < \delta$, so that $\boldsymbol{P}(t) > \mathbf{0}$ for $0 < t \leq h$ and some $h > 0$; $\mathbf{0}$ denotes the null matrix. Now, from the Chapmanov-Kolmogorov equation we have (component-wise) $\boldsymbol{P}(t+h) = \boldsymbol{P}(t)\boldsymbol{P}(h) > \mathbf{0}$ for $0 < t \leq h$, thus $\boldsymbol{P}(t) > \mathbf{0}, 0 < t \leq 2h$ etc., hence we can conclude that $\boldsymbol{P}(t) > \mathbf{0}, \forall\, t > 0$.

If some $q_{ij} = 0$, we replace this value by a positive value $\frac{1}{n}$ and adjust the respective diagonal element q_{ii} in such way the the final sum in the $i-$th row of $\boldsymbol{Q}$ remains zero. We denote this adjusted matrix by $\boldsymbol{Q}_n$; all its non-diagonal elements are positive and from the proof of the previous part, the same holds true for the elements of the matrix $\boldsymbol{P}_n(t) = e^{\boldsymbol{Q}_n t}$ and every $t > 0$. Obviously, due to continuity, $\boldsymbol{Q}_n \to \boldsymbol{Q}$ as $n \to \infty$. Then $\boldsymbol{P}_n(t) \to \boldsymbol{P}(t)$ and thus $p_{ij}(t) \geq 0$ for all $i, j \in S$ and all $t \geq 0$.

□

Example 3.4. *Model On-Off.* A machine can be either in an operating state or in a repair state. The operating time is a random variable exponentially distributed with the mean value $\frac{1}{\alpha}$, $\alpha > 0$. When a failure occurs, a repair is to be started immediately. The time of the repair is a random variable with the exponential distribution with the mean $\frac{1}{\beta}$, $\beta > 0$, independent of the operating time. When the machine is repaired it returns to the operating state. Define random variable X_t, that takes value 0, if at time t the machine is in the operating state and value 1, if it is in the repair. Then $\{X_t, t \geq 0\}$ is the continuous time Markov chain with state space $S = \{0, 1\}$ and the intensity matrix

$$\boldsymbol{Q} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}.$$

Now we will show how to obtain transition probabilities.

The backward differential equations for transition probabilities $p_{ij}(t)$ are

$$\begin{aligned} p'_{00}(t) &= -\alpha p_{00}(t) + \alpha p_{10}(t), \\ p'_{10}(t) &= \quad \beta p_{00}(t) - \beta p_{10}(t). \end{aligned} \tag{3.18}$$

Multiplying the first equation by β and the second one by α and then summing both equations we have

$$\beta p'_{00}(t) + \alpha p'_{10}(t) = 0,$$

and, after integrating,

$$\beta p_{00}(t) + \alpha p_{10}(t) = c$$

for a constant c. Condition $p_{ij}(0) = \delta_{ij}$ implies $c = \beta$ and

$$\beta p_{00}(t) + \alpha p_{10}(t) = \beta. \tag{3.19}$$

From (3.19) we compute $p_{10}(t)$ and insert into the first equation in (3.18) to get

$$p'_{00}(t) = \beta - (\alpha + \beta)p_{00}(t).$$

This is the differential equation for $p_{00}(t)$, the general solution of which is

$$p_{00}(t) = ce^{-(\alpha+\beta)t} + \frac{\beta}{\alpha + \beta}.$$

For $t = 0$ we have

$$p_{00}(0) = 1 = c + \frac{\beta}{\alpha + \beta},$$

thus

$$p_{00}(t) = \frac{\alpha}{\alpha + \beta}e^{-(\alpha+\beta)t} + \frac{\beta}{\alpha + \beta}.$$

The remaining probabilities then can be easily computed from (3.19).

The transition probabilities $p_{ij}(t)$ can be computed also directly from formula (3.17). It can be easily shown that the eigenvalues of $\boldsymbol{Q}$ are $\lambda_1 = 0, \lambda_2 = -(\alpha + \beta)$ and

$$\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^{-1},$$

where

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 \\ 0 & -(\alpha + \beta) \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 1 & \alpha \\ 1 & -\beta \end{pmatrix}.$$

Now according to (3.17)

$$\begin{aligned} \boldsymbol{P}(t) = e^{\boldsymbol{Q}t} &= \sum_{k=0}^{\infty} \frac{\boldsymbol{Q}^k t^k}{k!} = \sum_{k=0}^{\infty} \frac{t^k(\boldsymbol{B}\boldsymbol{A}\boldsymbol{B}^{-1})^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{t^k}{k!}\boldsymbol{B} \begin{pmatrix} 0^k & 0 \\ 0 & (-(\alpha + \beta))^k \end{pmatrix} \boldsymbol{B}^{-1} \\ &= \boldsymbol{B} \begin{pmatrix} 1 & 0 \\ 0 & e^{-(\alpha+\beta)t} \end{pmatrix} \boldsymbol{B}^{-1} \\ &= \frac{1}{\alpha + \beta} \begin{pmatrix} \beta + \alpha e^{-(\alpha+\beta)t} & \alpha - \alpha e^{-(\alpha+\beta)t} \\ \beta - \beta e^{-(\alpha+\beta)t} & \alpha + \beta e^{-(\alpha+\beta)t} \end{pmatrix}. \end{aligned} \tag{3.20}$$

To get (3.20) we can also use the Perron formula for the matrix $e^{\boldsymbol{Q}t}$ (Appendix B, Theorem B.6). We see that $\det(\lambda \boldsymbol{I} - \boldsymbol{Q}) = \lambda(\lambda+\alpha+\beta)$ and the adjoint matrix $\operatorname{adj}(\lambda \boldsymbol{I} - \boldsymbol{Q})$ is of the form

$$\operatorname{adj}(\lambda \boldsymbol{I} - \boldsymbol{Q}) = \begin{pmatrix} \lambda + \beta & \alpha \\ \beta & \lambda + \alpha \end{pmatrix}.$$

Further,

$$\psi_1(\lambda) = \frac{\det(\lambda \boldsymbol{I} - \boldsymbol{Q})}{\lambda - \lambda_1} = \lambda + \alpha + \beta,$$
$$\psi_2(\lambda) = \frac{\det(\lambda \boldsymbol{I} - \boldsymbol{Q})}{\lambda - \lambda_2} = \lambda.$$

Inserting into the Perron formula we get

$$\boldsymbol{P}(t) = e^{\boldsymbol{Q}t} = \frac{e^{\lambda_1 t}}{\psi_1(\lambda_1)} \operatorname{adj}(\lambda_1 \boldsymbol{I} - \boldsymbol{Q}) + \frac{e^{\lambda_2 t}}{\psi_2(\lambda_2)} \operatorname{adj}(\lambda_2 \boldsymbol{I} - \boldsymbol{Q}),$$

which is again (3.20).

Now, let us determine the absolute probabilities at time t. Suppose that the initial distribution is $p_0 = P(X_0 = 0) = 1,\ p_1 = P(X_0 = 1) = 0$. Then by (3.3)

$$p_0(t) = p_0 p_{00}(t) + p_1 p_{10}(t) = p_{00}(t),$$
$$p_1(t) = p_0 p_{01}(t) + p_1 p_{11}(t) = p_{01}(t).$$

In case of a denumerable time continuous Markov chains is the solution of the Kolmogorov equation more complex and for processes that are not regular the solution need not exist.Usually, we confine ourselves to computation of absolute probabilities

$p_j(t), j \in S$ with fixed initial state i, i.e., for the initial distribution $p_i(0) = 1, p_j(0) = 0, j \neq i$. Under the same conditions that hold for the forward Kolmogorov differential equations we can develop a system of differential equations for absolute probabilities: from (3.3) and (3.16) we have

$$p_j'(t) = -p_j(t)q_j + \sum_{k \neq j} p_k(t) q_{kj}, \quad j \in S$$

with the initial condition

$$p_i(0) = 1,\ p_j(0) = 0, \quad j \neq i.$$

Under this initial condition the absolute probabilities correspond to the $i-$th row of the matrix $\boldsymbol{P}(t)$. Some special cases will be solved later on.

3.3 Stationary and limit distribution

Similarly as for discrete time Markov chains we define the stationary distribution for continuous time Markov chains.

Definition 3.7. Let $\{X(t), t \geq 0\}$ be a continuous time Markov chain with the state space S and the transition probability matrices $\boldsymbol{P}(t), t \geq 0$. A vector $\boldsymbol{\eta} = \{\eta_i \geq 0, i \in S\}$ such that

$$\boldsymbol{\eta}^T \boldsymbol{P}(t) = \boldsymbol{\eta}^T, \quad t \in T \tag{3.21}$$

is called *invariant measure* of the process $\{X_t, t \geq 0\}$ on S. A probability vector $\boldsymbol{\pi}$ on S, that satisfies (3.21), is called *stationary distribution* of the given chain.

Theorem 3.12. *If the initial distribution of a homogeneous time continuous Markov chain $\{X_t, t \geq 0\}$ is stationary, then $\{X_t, t \geq 0\}$ is strictly stationary stochastic process, i.e., for any $k \in \mathbb{N}, 0 \leq t_1 \cdots < t_k$, for every $s > 0$ and $i_1, \ldots, i_k \in S$,*

$$P(X_{t_1} = i_1, \ldots, X_{t_k} = i_k) = P(X_{t_1+s} = i_1, \ldots, X_{t_k+s} = i_k).$$

Especially, for the absolute probabilities

$$p_j(t) = P(X_t = j) = \pi_j, \ j \in S, \ t \geq 0.$$

Proof. The proof is analogous to the proof of Theorem 2.24 if we use (3.2), (3.3) and the definition of the stationary distribution.

□

Definition 3.8. A probability distribution $\boldsymbol{a} = \{a_i, i \in S\}$ on S is called *limit distribution,* if for all $i, j \in S$

$$\lim_{t \to \infty} p_{ij}(t) = a_j.$$

Theorem 3.13. *If the limit distribution of a Markov chain exists it is a stationary distribution.*

Proof. Let $\boldsymbol{a} = \{a_i, i \in S\}$ be the limit distribution. Then for $t \geq 0, h \geq 0$ and positive integer N we get using the Chapman-Kolmogorov equation

$$p_{ij}(t+h) = \sum_{k \in S} p_{ik}(t) p_{kj}(h) \geq \sum_{k=0}^{N} p_{ik}(t) p_{kj}(h),$$

the limit as $t \to \infty$ gives

$$a_j \geq \sum_{k=0}^{N} a_k p_{kj}(h)$$

and limit as $N \to \infty$ gives

$$a_j \geq \sum_{k=0}^{\infty} a_k p_{kj}(h).$$

If for an index $j \in S$ the last inequality holds sharply, then by summing we realize that

$$\sum_{j \in S} a_j > \sum_{k \in S} a_k,$$

which is a contradiction, so that

$$a_j = \sum_{k \in S} a_k p_{kj}(h), h \geq 0,\ j \in S$$

is the stationary distribution.

□

Definition 3.9. We say that state j of a Markov chain $\{X_t, t \geq 0\}$ is accessible from state i, if

$$P_i(X_t = j) = p_{ij}(t) > 0 \quad \text{for some } t > 0, \quad i, j \in S.$$

Theorem 3.14. *The following relations between states of the continuous time Markov chain $\{X_t, t \geq 0\}$ and the respective embedded discrete time Markov chain $\{Y_n, n \in \mathbb{N}_0\}$ are equivalent:*

(1) *j is accessible from i in $\{X_t, t \geq 0\}$;*

(2) *j is accessible from i in $\{Y_n, n \in \mathbb{N}_0\}$;*

(3) *$q_{i_0 i_1} q_{i_1 i_2} \dots q_{i_{n-1} i_n} > 0$ for states $i_0 = i, i_n = j$ and some states $i_1, \dots, i_{n-1}$;*

(4) *$p_{ij}(t) > 0 \quad \forall\ t > 0$;*

(5) *$p_{ij}(t) > 0 \quad$ for some $t > 0$.*

Proof. Obviously (4) $\Rightarrow$ (5) $\Rightarrow$ (1).

If $p_{ij}(t) > 0$ for some $t > 0$, then $q_{ij}^{*(n)} > 0$ for some $n > 0$, where $q_{ij}^{*(n)}$ is the n-step transition probability in the chain $\{Y_n, n \in \mathbb{N}_0\}$, thus (1) $\Rightarrow$ (2).

If $q_{ij}^{*(n)} > 0$ for some $n > 0$, there exists a path from i to j in $\{Y_n\}$ such that $q_{ii_1}^* q_{i_1 i_2}^* \dots q_{i_{n-1}j}^* > 0$, so that also $q_{ii_1} q_{i_1 i_2} \dots q_{i_{n-1}j} > 0$, therefore (2) $\Rightarrow$ (3).

If $q_{ij} = p'_{ij}(0_+) > 0$, then $p_{ij}(s) > 0$ for all $0 < s \leq \delta$ where $\delta > 0$. Then for all $t \geq \delta$

$$p_{ij}(t) \geq p_{ij}(\delta) p_{jj}(t - \delta) > 0,$$

since

$$p_{jj}(t) = P_j(X_t = j) \geq e^{-q_j t} > 0 \text{ for all } t \geq 0,$$

thus, for $q_{ij} > 0$ it holds $p_{ij}(t) > 0$ for all $t > 0$. If (3) is valid, we can write

$$p_{ij}(t) \geq p_{ii_1}\left(\frac{t}{n}\right) p_{i_1 i_2}\left(\frac{t}{n}\right) \dots p_{i_{n-1}j}\left(\frac{t}{n}\right) > 0 \quad \forall\, t > 0,$$

hence, (3) $\Rightarrow$ (4).

□

From Theorem 3.14 it follows that the chain $\{X_t, t \geq 0\}$ is irreducible if and only if the respective embedded chain $\{Y_n, n \in \mathbb{N}_0\}$ is irreducible.

Definition 3.10. State j of the chain $\{X_t, t \geq 0\}$ is called *recurrent*, if either $q_j = 0$ (j is absorbing), or $q_j > 0$ and $P_j(\mathcal{T}_j(1) < \infty) = 1$, where

$$\mathcal{T}_j(1) = \inf\{t \geq J_1 : X_t = j\}$$

is the first return to state j.

State j is called *transient* if $q_j > 0$ and $P_j(\mathcal{T}_j(1) = \infty) > 0$.

A recurrent state j is called *positive,* if either $q_j = 0$, or $E_j \mathcal{T}_j(1) < \infty$.

It can be shown that a state that is recurrent in the embedded chain $\{Y_n\}$, is recurrent also in $\{X_t\}$ and vice versa; however, a state which is positive recurrent in the embedded chain $\{Y_n\}$, need not have the same property in $\{X_t\}$.

Theorem 3.15. *Let $\{X_t, t \geq 0\}$ be a continuous time Markov chain with the intensity matrix $\boldsymbol{Q}$ and the respective embedded chain $\{Y_n, n \in \mathbb{N}_0\}$, that is irreducible and all its states are recurrent. Then there exists an invariant measure $\boldsymbol{\eta}$ of the process $\{X_t, t \geq 0\}$,* *which is determined uniquely (up to a multiplicative constant) by the solution of the system*

$$\boldsymbol{\eta}^T \boldsymbol{Q} = \boldsymbol{0}^T \tag{3.22}$$

and $0 < \eta_j < \infty$ *for all* $j \in S$. *If* $\sum_{j \in S} \eta_j < \infty$, *then* $\boldsymbol{\pi} = \{\pi_j, j \in S\}$, *where*

$$\pi_j = \frac{\eta_j}{\sum_{k \in S} \eta_k},$$

is the stationary distribution of the process $\{X_t, t \geq 0\}$.

Proof. We will show the main steps of the proof, only. Let $\boldsymbol{\eta} = \{\eta_j \geq 0, j \in S\}$ be a solution of the equation $\boldsymbol{\eta}^T \boldsymbol{Q} = \boldsymbol{0}^T$. Considering this equation component-wise we have

$$\sum_{i \in S} \eta_i q_{ij} = \sum_{i \neq j} \eta_i q_{ij} + \eta_j q_{jj} = 0, \quad j \in S,$$

or

$$\sum_{i \neq j} \eta_i q_{ij} = \eta_j q_j, \quad j \in S.$$

Using (3.12) we get

$$\sum_{i \in S} \eta_i q_i q^*_{ij} = \eta_j q_j, \quad j \in S,$$

which means that $\{\eta_j q_j, j \in S\}$ is an invariant measure of the embedded discrete time Markov chain with the transition probability matrix $\boldsymbol{Q}^* = \{q^*_{ij}, i, j \in S\}$. By the assumption, this chain is irreducible with recurrent states, thus, up to a multiplicative constant,

$$0 < \eta_j q_j = E_i \sum_{n=0}^{\tau_i(1)-1} I(Y_n = j) < \infty$$

for a fixed $i \in S$, where $\tau_i(1)$ is the first passage time to i in the chain $\{Y_n, n \in \mathbb{N}_0\}$.

Further, we show that $\eta_j = \mu^i_j$, where

$$\mu^i_j = E_i \int_0^{\mathcal{T}_i(1)} I(X_t = j) dt$$

is the mean holding time in state j between two returns of the chain $\{X_t, t \geq 0\}$ to state i. Namely,

$$\mu^i_j = \int_0^\infty P_i(X_t = j, \mathcal{T}_i(1) > t) dt = \int_0^\infty \sum_{n=0}^\infty P_i(Y_n = j, J_n \leq t < J_{n+1}, \mathcal{T}_i(1) > t) dt.$$

Notice that $\mathcal{T}_i(1), \tau_i(1)$ are stopping times such that

$$[\mathcal{T}_i(1) > J_n] \Longleftrightarrow [\tau_i(1) > n],$$

and using conditioning and the Markov property we get

$$\begin{aligned}
\mu_j^i &= \sum_{n=0}^{\infty} \int_0^{\infty} P_i(J_n \le t < J_{n+1} | Y_n = j) P_i(Y_n = j, \tau_i(1) > n) dt \\
&= \sum_{n=0}^{\infty} E_i(T_{n+1} | Y_n = j) P_i(Y_n = j, \tau_i(1) > n) \\
&= \frac{1}{q_j} E_i \sum_{n=0}^{\tau_i(1)-1} I(X_n = j) = \eta_j.
\end{aligned}$$

Now, we need to show that $\boldsymbol{\eta}$ is an invariant measure of the process $\{X_t, t \ge 0\}$. Since

$$E_i \int_0^s I(X_t = j) dt = E_i \int_{\mathcal{T}_i(1)}^{s+\mathcal{T}_i(1)} (X_t = j) dt,$$

we have

$$\begin{aligned}
\eta_j = \mu_j^i &= E_i \int_0^s I(X_t = j) dt + E_i \int_s^{\mathcal{T}_i(1)} I(X_t = j) dt = E_i \int_s^{\mathcal{T}_i(1)+s} I(X_t = j) dt \\
&= \int_0^{\infty} P_i(X_{t+s} = j, \mathcal{T}_i(1) > t) dt \\
&= \int_0^{\infty} \sum_{k=0}^{\infty} P_i(X_{t+s} = j | X_t = k, \mathcal{T}_i(1) > t) P_i(X_t = k, \mathcal{T}_i(1) > t) dt,
\end{aligned}$$

from here, again using the Markov property

$$\begin{aligned}
\eta_j &= \sum_{k=0}^{\infty} \int_0^{\infty} P_i(X_{t+s} = j | X_t = k) P_i(X_t = k, \mathcal{T}_i(1) > t) dt \\
&= \sum_{k=0}^{\infty} p_{kj}(s) E_i \int_0^{\mathcal{T}_i(1)} I(X_t = k) dt \\
&= \sum_{k=0}^{\infty} p_{kj}(s) \mu_k^i = \sum_{k=0}^{\infty} p_{kj}(s) \eta_k,
\end{aligned}$$

thus, $\boldsymbol{\eta}$ is the invariant measure. If $\sum_{j \in S} \eta_j < \infty$, there exists a stationary distribution of the chain $\{X_t, t \ge 0\}$ (we can set $\pi_j = \eta_j / \sum_{k \in S} \eta_k$ for $j \in S$).

□

Theorem 3.16. *Let $\{X_t, t \geq 0\}$ be a continuous time Markov chain with the intensity matrix $\boldsymbol{Q}$ and the embedded chain $\{Y_n, n \in \mathbb{N}_0\}$, which is irreducible and all the states are recurrent. Let $\boldsymbol{\pi} = \{\pi_j, j \in S\}$, where $\pi_j > 0$ for all $j \in S$ and $\sum_{j\in S} \pi_j = 1$, be a solution to*

$$\boldsymbol{\pi}^T \boldsymbol{Q} = \boldsymbol{0}^T.$$

Then the transition respectively the absolute probabilities in the chain $\{X_t, t \geq 0\}$ satisfy

$$\lim_{t\to\infty} p_{ij}(t) = \pi_j \text{ for all } i, j \in S, \tag{3.23}$$

$$\lim_{t\to\infty} p_j(t) = \pi_j \text{ for all } j \in S. \tag{3.24}$$

Proof. According to Theorem 3.15, $\boldsymbol{\pi}$ is the stationary distribution of $\{X_t, t \geq 0\}$, and thus $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \boldsymbol{P}(t)$ for all $t \geq 0$. Since the embedded chain is irreducible, we have from Theorem 3.14 that $p_{ij}(t) > 0$ for all $t > 0$ and all $i, j \in S$.

Now, let us consider discrete times $h, 2h, 3h, \ldots$ for some $h > 0$ and a sequence of random variables $\{Z_n, n \in \mathbb{N}_0\}$ defined by $Z_n = X_{nh}$. Obviously it holds

$$P(Z_n = j|Z_{n-1} = i, \ldots, Z_0 = i_0) = P(X_{nh} = j|X_{(n-1)h} = i, \ldots, X_0 = i_0) = p_{ij}(h),$$

so that $\{Z_n, n \in \mathbb{N}_0\}$ is a discrete time Markov chain with the transition probability matrix $\boldsymbol{P}(h)$ and $\boldsymbol{\pi}$ is the stationary distribution in the chain $\{Z_n\}$. Please notice that this is not the embedded Markov chain of jumps defined by Definition 3.6. According to Theorem 2.25 the elements of $\boldsymbol{P}^n(h)$ satisfy

$$\lim_{n\to\infty} p_{ij}^{(n)}(h) = \lim_{n\to\infty} p_{ij}(nh) = \pi_j \text{ for all } i, j \in S,$$

thus,

$$\forall\, \varepsilon > 0 \;\; \exists N \in \mathbb{N} \;\; \forall\, n > N \;\; |p_{ij}(nh) - \pi_j| < \varepsilon.$$

Further we have

$$|p_{ij}(t) - \pi_j| \leq |p_{ij}(t) - p_{ij}(nh)| + |p_{ij}(nh) - \pi_j|$$

and due to uniform continuity of $p_{ij}(t)$ (Theorem 3.1) it holds $|p_{ij}(t) - p_{ij}(nh)| < \varepsilon$ for all $nh \leq t \leq (n+1)h$ and sufficiently small h. From here the assertion (3.23) follows. The assertion (3.24) follows immediately if we apply limit as $t \to \infty$ to

$$p_j(t) = \sum_{i\in S} p_i(0) p_{ij}(t), \quad j \in S.$$

□

Example 3.5. *(Example 3.4 continued)* Consider again the continuous time Markov chain with states $0, 1$ and the intensity matrix

$$\boldsymbol{Q} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}.$$

In Example 3.4 we have computed the transition probabilities matrices

$$\boldsymbol{P}(t) = \frac{1}{\alpha+\beta}\begin{pmatrix} \beta + \alpha e^{-(\alpha+\beta)t} & \alpha - \alpha e^{-(\alpha+\beta)t} \\ \beta - \beta e^{-(\alpha+\beta)t} & \alpha + \beta e^{-(\alpha+\beta)t} \end{pmatrix}.$$

We can see that

$$\lim_{t\to\infty} \boldsymbol{P}(t) = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}.$$

The componenents of the vector

$$\boldsymbol{\pi} = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)^T$$

give the stationary distribution of the considered Markov chain. We obtain the same result by solving the system of equations (3.22), that is

$$-\pi_0\alpha + \pi_1\beta = 0,$$
$$\pi_0\alpha - \pi_1\beta = 0\ .$$

The solution that satisfies condition $\pi_0 + \pi_1 = 1$, is again

$$\pi_0 = \frac{\beta}{\alpha+\beta}, \quad \pi_1 = \frac{\alpha}{\alpha+\beta}\ .$$

Example 3.6. *Telephone exchange.* Consider a telephone exchange with N channels. Assume that in the interval $(t, t+h]$ a call arrives with the probability $\lambda h + o(h),\ \lambda > 0$, which is the same for all $t \geq 0$. The calls arrive independently each other. The probability that in $(t, t+h]$ two or more calls arrive is $o(h)$, the probability of no call is $1 - \lambda h + o(h)$. If all N channels are busy, a new arriving call disappears. Epochs for which the calls last are independent random variables, independent of the call arrivals.

Further let us suppose that the epoch in which a call is lasting is a random variable T exponentially distributed with the mean value $\frac{1}{\mu}$, $\mu > 0$. The probability that a call that lasts at time t, will be finished in $(t, t+h]$ is

$$P(t < T \leq t+h | T > t) = \frac{P(t < T \leq t+h)}{P(T > t)} = 1 - e^{-\mu h} = \mu h + o(h).$$

The probability that a call that continues at time t, will not finish in $(t, t+h]$, is $1 - \mu h + o(h)$.

Let X_t be the number of busy channels at time t; then $\{X_t, t \geq 0\}$ is a continuous time Markov chain with the state space $S = \{0, 1, \ldots, N\}$. For the transition probabilities $p_{ij}(t, t+h) = p_{ij}(h)$ we have

$$p_{j,j+1}(h) = [\lambda h + o(h)][1 - \mu h + o(h)] + o(h) = \lambda h + o(h),$$

similarly,

$$p_{j,j-1}(h) = \binom{j}{1}[\mu h + o(h)][1 - \mu(h) + o(h)]^{j-1}[1 - \lambda h + o(h)] + o(h) = j\mu h + o(h)$$

and

$$\begin{aligned} p_{j,j+k}(h) &= o(h), \quad 2 \leq k \leq N - j, \\ p_{j,j-k}(h) &= o(h), \quad 2 \leq k \leq j, \\ p_{jj}(h) &= 1 - (\lambda + j\mu)h + o(h), \quad 1 \leq j \leq N-1, \\ p_{00}(h) &= 1 - \lambda h + o(h), \\ p_{NN}(h) &= 1 - N\mu h + o(h). \end{aligned}$$

From here we compute intensities

$$\begin{aligned} q_{j,j+1} &= \lambda, & 0 \leq j \leq N-1, \\ q_{j,j-1} &= j\mu, & 1 \leq j \leq N, \\ q_j &= \lambda + j\mu, & 1 \leq j \leq N-1, \\ q_0 &= \lambda, \\ q_N &= N\mu, \\ q_{ij} &= 0 \quad \text{otherwise} . \end{aligned}$$

The intensity matrix is

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \ldots & 0 & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 & \ldots & 0 & 0 & 0 \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & (N-1)\mu & -(\lambda+(N-1)\mu) & \lambda \\ 0 & 0 & 0 & 0 & \ldots & 0 & N\mu & -N\mu \end{pmatrix}.$$

It can be easily checked that the embedded chain $\{Y_n, n \in \mathbb{N}_0\}$ is irreducible and the conditions of Theorem 3.16 are fulfilled. The system of equations $\boldsymbol{\pi}^T \boldsymbol{Q} = \boldsymbol{0}^T$ writes

$$\begin{aligned} -\lambda\pi_0 + \mu\pi_1 &= 0, \\ \lambda\pi_{j-1} - (\lambda + j\mu)\pi_j + (j+1)\mu\pi_{j+1} &= 0, \quad 1 \leq j \leq N-1, \\ \lambda\pi_{N-1} - N\mu\pi_N &= 0. \end{aligned} \tag{3.25}$$

We can solve these equations one after the other one; we obtain the recursive relation

$$\pi_j = \pi_0 \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!}, \quad 1 \le j \le N. \tag{3.26}$$

From here together with the condition $\sum_{j=0}^N \pi_j = 1$ we have

$$\pi_j = \frac{\rho^j}{j!} \left(\sum_{k=0}^N \frac{\rho^k}{k!}\right)^{-1}, \quad 0 \le j \le N,$$

where $\rho = \frac{\lambda}{\mu}$.

Notice that relation (3.26) can be obtained also by a substitution

$$K_j = j\mu\pi_j - \lambda\pi_{j-1},\ 1 \le j \le N.$$

Then (3.25) takes the form

$$\begin{aligned} K_1 &= 0, \\ K_{j+1} - K_j &= 0, \quad 1 \le j \le N-1, \\ K_N &= 0, \end{aligned}$$

hence $K_j = 0,\ 1 \le j \le N$, and from here we again get (3.26).

Example 3.7. In a factory with uninterrupted traffic, N automatic machines are working that are served by $r < N$ repairmen. Any machine can break down independently of its previous state as well as of the states of other machines. The machine that breaks down is taken into repair if any servisman is idle; if all r servicemen are busy, the machine is waiting for the repair. We assume that the repair of one machine needs the service of only one serviceman, and the servicemen are working independently each other. The repair epochs are random variables with the same exponential distribution.

Assume that any machine that is working at time t breaks down at $(t, t+h]$ with probability $\lambda h + o(h),\ \lambda > 0$, which is the same for all machines. A machine that is in the service at time t, is during $(t, t+h]$ repaired and returned to the working state with probability $\mu h + o(h),\ \mu > 0$.

Let X_t be the number of machines that at time t do not work. Under our assumptions, $\{X_t, t \ge 0\}$ is a homogeneous continuous time Markov chain with the state space $S = \{0, 1, \dots, N\}$.

Similarly as in the previous example we get

$$\begin{aligned} p_{j,j+1}(h) &= (N-j)\lambda h + o(h), && 0 \le j < N, \\ p_{j,j-1}(h) &= j\mu h + o(h), && 1 \le j \le r, \\ p_{j,j-1}(h) &= r\mu h + o(h), && r < j \le N, \\ p_{j,j+k}(h) &= o(h), && 2 \le k \le N-j, \\ p_{j,j-k}(h) &= o(h), && 2 \le k \le j \end{aligned}$$

and $p_{jj}(h)$ can be derived in the same way. The transition intensities are

$$\begin{aligned} q_{j,j+1} &= (N-j)\lambda, && 0 \le j < N, \\ q_{j,j-1} &= j\mu, && 1 \le j \le r, \\ q_{j,j-1} &= r\mu, && r < j \le N, \\ q_{jk} &= 0, && j \ne k \text{ otherwise} \end{aligned}$$

and the total intensities are

$$\begin{aligned} q_0 &= N\lambda, \\ q_j &= (N-j)\lambda + j\mu, \quad 1 \le j \le r, \\ q_j &= (N-j)\lambda + r\mu, \quad r < j < N, \\ q_N &= r\mu. \end{aligned}$$

Let us compute limit probabilities. By using Theorem 3.16 we get the system of equations

$$\begin{aligned} -N\lambda\pi_0 + \mu\pi_1 &= 0, \\ (N-j+1)\lambda\pi_{j-1} - ((N-j)\lambda + j\mu)\pi_j + (j+1)\mu\pi_{j+1} &= 0, \quad 1 \le j < r, \\ (N-j+1)\lambda\pi_{j-1} - ((N-j)\lambda + r\mu)\pi_j + r\mu\pi_{j+1} &= 0, \quad r \le j < N, \\ \lambda\pi_{N-1} - r\mu\pi_N &= 0. \end{aligned}$$

Put

$$K_j = j\mu\pi_j - (N-j+1)\lambda\pi_{j-1}, \quad K_j^* = r\mu\pi_j - (N-j+1)\lambda\pi_{j-1}.$$

Then the system of equations can be rewritten as

$$\begin{aligned} K_1 &= 0, \\ K_{j+1} - K_j &= 0, \quad 1 \le j < r, \\ K_{j+1}^* - K_j^* &= 0, \quad r \le j < N, \\ K_N^* &= 0. \end{aligned}$$

From here together with $K_r = K_r^*$ we have $K_j = 0,\ 1 \le j \le r, \quad K_j^* = 0,\ r \le j \le N$. We obtain recursive relations

$$\begin{aligned} \pi_j &= \frac{(N-j+1)\lambda}{j\mu}\pi_{j-1}, \quad 1 \le j < r, \\ \pi_j &= \frac{(N-j+1)\lambda}{r\mu}\pi_{j-1}, \quad r \le j \le N, \end{aligned}$$

from which we develop that

$$\pi_j = \binom{N}{j}\left(\frac{\lambda}{\mu}\right)^j \pi_0, \qquad 1 \le j < r,$$
$$\pi_j = \frac{N(N-1)\dots(N-j+1)}{r!r^{j-r}}\left(\frac{\lambda}{\mu}\right)^j \pi_0, \quad r \le j \le N.$$

The value of π_0 follows from the condition $\sum_{j=0}^{N} \pi_j = 1$.

3.4 Poisson process

In Example 3.1 we have considered the process $\{X_t, t \ge 0\}$ of integer-valued random variables with independent and stationary increments $X_{t+h} - X_t$ (the number of events in the interval $(t, t+h]$). From the assumptions stated there it also holds

$$P(X_{t+h} - X_t = 1) = \lambda h + o(h),$$
$$P(X_{t+h} - X_t = 0) = 1 - \lambda h + o(h),$$
$$P(X_{t+h} - X_t \ge 2) = o(h)$$

uniformly for all t.

We know that $\{X_t, t \ge 0\}$ is a continuous time Markov chain with the state space $S = \{0, 1, \dots\}$, the initial distribution $p_0(0) = P(X_0 = 0) = 1,\ p_j(0) = 0,\ j \ne 0$ and the transient intensities $q_{i,i+1} = \lambda, q_i = -q_{ii} = \lambda, q_{ij} = 0$ otherwise. The transition probabilities $p_{ij}(t)$ can be computed by solving the system of the Kolmogorov forward equations (3.16). If we confine ourselves to the absolute probabilities $p_j(t), j \in S$ only and the initial distribution $p_i = p_i(0) = 1, p_j = p_j(0) = 0,\ j \ne i$, we can solve a system of differential equations that correspond to the $i-$th row of the matrix $\boldsymbol{P}(t)$, i.e., to solve the system of the differential equations

$$p'_j(t) = -p_j(t)q_j + \sum_{k \ne j} p_k(t)q_{kj}, \quad j \in S$$

with the initial condition

$$p_i(0) = 1,\ p_j(0) = 0, \quad j \ne i,$$

in our case,

$$p'_0(t) = -\lambda p_0(t), \tag{3.27}$$
$$p'_j(t) = -\lambda p_j(t) + \lambda p_{j-1}(t), \quad 1 \le j < \infty$$

with the initial condition

$$p_0(0) = 1, \quad p_j(0) = 0, \ j > 0. \tag{3.28}$$

The system (3.27) is a system of linear differential equations. We can solve it either one by one, but we can use the method of the generating function.

Let us consider the probability generating function of the distribution $\{p_j(t), j \in \mathbb{N}_0\}$,

$$\Pi(s,t) = \sum_{j=0}^{\infty} p_j(t) s^j,$$

as a function of variables s, t. If we multiply the $j-$th equation of (3.27) by s^j and then add all such multiplied equations we get

$$\sum_{j=0}^{\infty} p_j'(t) s^j = -\lambda \sum_{j=0}^{\infty} s^j p_j(t) + \lambda \sum_{j=1}^{\infty} p_{j-1}(t) s^j = -\lambda \sum_{j=0}^{\infty} s^j p_j(t) + \lambda s \ \sum_{j=0}^{\infty} p_j(t) s^j.$$

This expression can be written in terms of the probability generating function Π as

$$\frac{\partial \Pi(s,t)}{\partial t} = -\lambda \Pi(s,t) + \lambda s \ \Pi(s,t) = -\lambda(1-s)\Pi(s,t) \tag{3.29}$$

with the initial condition

$$\Pi(s,0) = 1. \tag{3.30}$$

For fixed s, (3.29) is an ordinary linear differential equation for the variable t; the general solution of which is

$$\Pi(s,t) = C(s) e^{-\lambda t(1-s)},$$

where $C(s)$ is a constant dependent on s, only. From the initial condition (3.30) we immediately have $C(s) = 1$, thus,

$$\Pi(s,t) = e^{-\lambda t + \lambda s t} = e^{-\lambda t} \sum_{j=0}^{\infty} \frac{(\lambda t)^j}{j!} s^j.$$

From here, we can conclude that the absolute distribution is given by probabilities

$$p_j(t) = \frac{e^{-\lambda t}(\lambda t)^j}{j!}, \quad 0 \le j < \infty, \quad t > 0$$

which is the Poisson distribution with the parameter λt.

Since the number of events that occur in an arbitrary interval $(s, s+t]$ depends only on the length of this interval, it follows from here, that the increments $X_{s+t} - X_s$ have the Poisson distribution with the parameter λt for all $s, t > 0$.

Remark. A process $\{X_t, t \geq 0\}$ of integer-valued random variables that represent numbers of some events that could occur in an interval $[0, t]$ is called *counting process.* The Poisson process with intensity λ is a counting process with independent and stationary increments such that $X_t = 0$ and for $t > 0$, X_t has the Poisson distribution with the parameter λt. Other properties of the Poisson process can be found, e.g., in Resnick (1992), Chap. 4.

Let $T_j, j = 1, 2, \dots$ be time between the occurrence of events $j-1$ and j, i.e., the holding time at state $j-1$. According to Theorem 3.5, T_j are random variables with the exponential distribution with the same parameter λ, i.e., with the mean value $\frac{1}{\lambda}$. We will show that T_j are mutually independent.

For T_1 and T_2 we have

$$P(T_1 > x, T_2 > y) = \int_x^\infty P(T_2 > y | T_1 = u\ )\lambda e^{-\lambda u}\, du\ ,$$

further, by Theorem 3.4

$$P(T_2 > y | T_1 = u\ ) = P(X_t = 1, u\ \leq t \leq y + u\ | X_u = 1) = e^{-\lambda y},$$

thus,

$$P(T_1 > x, T_2 > y) = e^{-\lambda y} \int_x^\infty \lambda e^{-\lambda u} du = e^{-\lambda x} e^{-\lambda y} = P(T_1 > x) P(T_2 > y);$$

from here the independence of T_1, T_2 follows. Similarly, we can prove the independence of random variables $T_1, T_2, \dots, T_n,\ n \geq 2$. The result is in accordance with the assertion of Theorem 3.8.

The embedded discrete time Markov chain $\{Y_n, n \in \mathbb{N}_0\}$ of the Poisson process has the transition probability matrix $\boldsymbol{Q}^*$ with the elements $q^*_{ii} = 0,\ q^*_{i,i+1} = 1,\ q^*_{ij} = 0$ otherwise. The Poisson process is regular, since $q_i = \lambda\ \forall.$, Thus, by Theorem 3.9

$$\sum_{i=0}^\infty \frac{1}{q_{Y_i}} = \sum_{i=0}^\infty \frac{1}{\lambda} = \infty$$

holds with probability one.

Now, let us consider the distribution of the random variable $W_n = S_1 + \cdots + S_n$, (the waiting time for the occurrence of the $n-$th event of a Poisson process). Since $S_1, \dots, S_n$ are independent random variables having the same exponential distribution, the distribution of W_n is given by the density

$$w_n(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{n-1}}{(n-1)!}, & x \geq 0, \\ 0 & \text{elsewhere} \end{cases}$$

(Erlang distribution of order n).

Example 3.8. Buses of lines No. 1 and No. 2 arrive to a bus-stop. The arrivals of buses of both lines are events of the Poisson processes with intensities λ_1 a λ_2, respectively, these processes are independent. The mean number of the buses of line No. 1 that arrive to the bus-stop during a time interval of length t is $\lambda_1 t$, the mean number of all buses arriving during this interval is $(\lambda_1 + \lambda_2)t$. Waiting time S for an arrival of a bus of Line 1 has the exponential distribution with the mean value $\frac{1}{\lambda_1}$, waiting time T for an arrival of Line 2 has the exponential distribution with the mean value $\frac{1}{\lambda_2}$. The waiting time for the arrival of the next (second in the order) arrival of Line 1 has the Erlang distribution of order 2, with parameter λ_1, and the mean value $\frac{2}{\lambda_1}$.

The probability that the first bus that will arrive is bus of Line 2, is

$$P(T < S\ ) = \int_0^\infty \left(\int_0^s \lambda_2 e^{-\lambda_2 t} dt \right) \lambda_1 e^{-\lambda_1 s} ds = \int_0^\infty \lambda_1 e^{-\lambda_1 s}\ (1 - e^{-\lambda_2 s}) ds = \frac{\lambda_2}{\lambda_1 + \lambda_2}.$$

Since the remaining time of the arrival of a bus of Line 1 has the exponential distribution with parameter λ_1 and times between arrivals are independent, the probability that the first two buses that will arrive to the bus-stop will be those of Line 2 is $(\frac{\lambda_2}{\lambda_1+\lambda_2})^2$.

3.5 Linear birth process (Yule process)

Consider a population of individuals that can reproduce but cannot cease to exist. Assume that during the interval $(t, t+h]$ each individual can give a birth of a new individual with the probability $\lambda h + o(h)$ independently of the behaviour of other individuals.

Let X_t denotes the number of individuals that are in the population at time t; it is an element of a continuous time Markov chain with the state space $S \subset \mathbb{N}_0$; for simplicity assume that $S = \{1, 2, \dots\}$. The transition probabilities are

$$\begin{aligned} p_{j,j+1}(h) &= \binom{j}{1} (\lambda h + o(h))(1 - \lambda h + o(h))^{j-1} = j\lambda h + o(h), \\ p_{j,j+k}(h) &= o(h), \quad k \geq 2, \\ p_{jj}(h) &= (1 - \lambda h + o(h))^j = 1 - j\lambda h + o(h); \end{aligned}$$

other probabilities are zeros. The intensities are

$$q_{j,j+1} = j\lambda,\ q_j = j\lambda \text{ for } 1 \leq j < \infty,\ q_{jk} = 0 \text{ otherwise},$$

thus the intensity matrix is

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ 0 & -2\lambda & 2\lambda & 0 & \dots \\ 0 & 0 & -3\lambda & 3\lambda & \dots \\ 0 & 0 & 0 & -4\lambda & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Assume that at the beginning, the population consists only of one individual, i.e., the initial distribution is $p_1(0) = 1,\ p_j(0) = 0,\ j > 1$. If we confine ourselves to the absolute probabilities, we realize that they satisfy the system of differential equations

$$\begin{aligned} p_1'(t) &= -\lambda p_1(t), \\ p_j'(t) &= -\lambda j p_j(t) + \lambda(j-1)p_{j-1}(t), \quad j > 1 \end{aligned} \tag{3.31}$$

with the initial condition $p_1(0) = 1$. System (3.31) can again be solved by the method of generating function: multiplying the $j-$th equation by s^j and adding all the equations we get

$$\begin{aligned} \sum_{j=1}^{\infty} p_j'(t)s^j &= -\lambda \sum_{j=1}^{\infty} j p_j(t) s^j + \lambda \sum_{j=2}^{\infty} (j-1) p_{j-1}(t) s^j \\ &= -\lambda s \sum_{j=1}^{\infty} j p_j(t) s^{j-1} + \lambda s^2 \sum_{j=1}^{\infty} j p_j(t) s^{j-1}, \end{aligned}$$

or, in terms of the probability generating function Π of the sequence $\{p_j(t)\}$,

$$\frac{\partial \Pi(s,t)}{\partial t} = -\lambda s \frac{\partial \Pi(s,t)}{\partial s} + \lambda s^2 \frac{\partial \Pi(s,t)}{\partial s},$$

equivalently,

$$\lambda s(1-s)\frac{\partial \Pi(s,t)}{\partial s} + \frac{\partial \Pi(s,t)}{\partial t} = 0 \tag{3.32}$$

which is a partial difference equation with the initial condition

$$\Pi(0,s) = s. \tag{3.33}$$

A solution of equation (3.32) is given by Theorem B.7 in the Appendix B. According to this theorem, we first solve the auxiliary ordinary differential equation

$$\frac{ds}{\lambda s\ (1-s)} = dt.$$

Solving it we have for $0 < s < 1$

$$\ln s\ - \ln(1-s) = \lambda t + C,$$

or

$$\frac{s}{1-s} = C_1 e^{\lambda t}.$$

The first integral of the auxiliary differential equation is $\frac{s}{1-s}e^{-\lambda t} = C_1$ and according to Theorem B.7 the general solution of equation (3.32) has form

$$\Pi(s,t) = F\left(\frac{s}{1-s}e^{-\lambda t}\right),$$

where F is a differentiable function. The initial condition (3.33) implies that

$$F\left(\frac{s}{1-s}\right) = s.$$

If we set $\frac{s}{1-s} = x$, we obtain

$$F(x) = \frac{x}{1+x},$$

thus

$$\Pi(s,t) = \frac{s}{s - s\; e^{\lambda t} + e^{\lambda t}} = \frac{s e^{-\lambda t}}{1 - s(1 - e^{-\lambda t})}.$$

Expanding the last expression into a power series we get

$$\Pi(s,t) = \sum_{k=0}^{\infty} s e^{-\lambda t} \left[s(1 - e^{-\lambda t})\right]^k = \sum_{k=1}^{\infty} e^{-\lambda t} \left(1 - e^{-\lambda t}\right)^{k-1} s^k,$$

and from here we can conclude that

$$P(X_t = k) = p_k(t) = e^{-\lambda t}\left(1 - e^{-\lambda t}\right)^{k-1}, \quad k = 1, 2, \ldots.$$

We easily realize that $\sum_{k=1}^{\infty} p_k(t) = 1$ for every $t \geq 0$ (the geometric distribution). The mean size of the population at time t is $EX_t = e^{\lambda t}$.

Similarly we can show that under the initial condition $p_{i_0}(0) = 1$ for some $i_0 > 1$ the generating function of absolute probabilities is given by

$$\Pi(s,t) = \left[\frac{s e^{-\lambda t}}{1 - s(1 - e^{-\lambda t})}\right]^{i_0},$$

(negative-binomial distribution).

The process is regular since under the initial condition $X_0 = i_0 \geq 1$ the states of the embedded chain $\{Y_n\}$ satisfy $P_{i_0}(Y_i = i + i_0) = 1, i = 1, 2, \ldots$ and thus

$$\sum_{i=1}^{\infty} \frac{1}{q_{Y_i}} = \sum_{i=1}^{\infty} \frac{1}{\lambda(i + i_0)} = \infty$$

holds with probability 1.

3.6 General birth process

Consider again a population of individuals that can reproduce only. The size of the population at time $t \geq 0$ is described by a time continuous Markov chain $\{X_t, t \geq 0\}$ with a countable state space, that is defined by the initial distribution $p_{i_0}(0) = 1,\ p_j(0) = 0,\ j > i_0\ (i_0 \geq 0)$ and the transition intensity matrix

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda_{i_0} & \lambda_{i_0} & 0 & 0 & \dots \\ 0 & -\lambda_{i_0+1} & \lambda_{i_0+1} & 0 & \dots \\ 0 & 0 & -\lambda_{i_0+2} & \lambda_{i_0+2} & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

where the intensities $\lambda_j > 0$ express a general dependence on state j of the population.

Absolute probabilities $P(X_t = j) = p_j(t)$ satisfy a system of differential equations

$$\begin{aligned} p'_{i_0}(t) &= -\lambda_{i_0} p_{i_0}(t), \\ p'_j(t) &= \lambda_{j-1} p_{j-1}(t) - \lambda_j p_j(t), \quad j > i_0 \end{aligned}$$

with the initial condition $p_{i_0}(0) = 1$.

Solving these equations sequentially and using the initial condition we get

$$\begin{aligned} p_{i_0}(t) &= e^{-\lambda_{i_0} t}, \\ p_j(t) &= \lambda_{j-1} e^{-\lambda_j t} \int_0^t e^{\lambda_j s} p_{j-1}(s) ds, \quad j > i_0. \end{aligned}$$

The process is regular if and only if

$$\sum_{j=i_0}^{\infty} \frac{1}{\lambda_j} = \infty,$$

since for the corresponding embedded chain $\{Y_n\}$ we have $P_{i_0}(Y_n = n + i_0) = 1$ for every $i_0 > 0$ and further we can proceed as in Theorem 3.9.

3.7 Linear birth and death process

Consider again the process described in Example 3.2.

If the size of the population at time t is X_t, then it follows from the assumptions given in Example 3.2 that $\{X_t, t > 0\}$ is the continuous time Markov chain with the state space $S = \{0, 1, \ldots, \}$ and intensities

$$\begin{aligned} q_{j,j+1} &= j\lambda, \quad 0 \le j < \infty, \\ q_{j,j-1} &= j\mu, \quad 1 \le j < \infty, \\ q_{jk} &= 0 \qquad \text{jinak}, \\ q_j &= j(\lambda + \mu), \quad 0 \le j < \infty \end{aligned}$$

The absolute probabilities $p_j(t) = P(X_t = j)$ satisfy the system of differential equations

$$\begin{aligned} &p_0'(t) = \mu p_1(t), \\ &p_j'(t) = (j-1)\lambda p_{j-1}(t) - j(\mu + \lambda)p_j(t) + (j+1)\mu p_{j+1}(t), \ j = 1, 2, \ldots \end{aligned}$$

Assume that $p_1(0) = 1$. Then by using the method of generating function we get the equation

$$\sum_{j=0}^{\infty} p_j'(t)s^j = \sum_{j=0}^{\infty}(j+1)\mu p_{j+1}(t)s^j + \sum_{j=1}^{\infty}(j-1)\lambda p_{j-1}(t)s^j - \sum_{j=1}^{\infty} j(\lambda+\mu)p_j(t)s^j,$$

or

$$\begin{aligned} \frac{\partial \Pi(s,t)}{\partial t} &= \mu\frac{\partial \Pi(s,t)}{\partial s} + \lambda s^2\frac{\partial \Pi(s,t)}{\partial s} - s(\lambda+\mu)\frac{\partial \Pi(s,t)}{\partial s} \\ &= [\mu + \lambda s^2 - s(\lambda+\mu)]\frac{\partial \Pi(s,t)}{\partial s} = (\mu - \lambda s)(1-s)\frac{\partial \Pi(s,t)}{\partial s} \end{aligned} \tag{3.34}$$

with the initial condition

$$\Pi(s,0) = s\ . \tag{3.35}$$

For $\lambda = \mu$ (3.34) is of the form

$$\lambda(1-s)^2\frac{\partial \Pi(s,t)}{\partial s} - \frac{\partial \Pi(s,t)}{\partial t} = 0. \tag{3.36}$$

The auxiliary ordinary differential equation is

$$\frac{ds}{\lambda(1-s)^2} = -dt$$

and its first integral is

$$\frac{1}{1-s} + \lambda t = C,$$

which means that the general solution of equation (3.36) is

$$\Pi(s,t) = F\left(\lambda t + \frac{1}{1-s}\right)$$

for a differentiable function F. It can be determine from the initial condition (3.35). We get

$$\Pi(s,0) = s \ = F\left(\frac{1}{1-s}\right)$$

and using the substitution $x = \frac{1}{1-s}$, we obtain that

$$F(x) = \frac{x-1}{x}.$$

The solution we were looking for is

$$\begin{aligned}\Pi(s,t) &= \frac{\lambda t + \frac{1}{1-s} - 1}{\lambda t + \frac{1}{1-s}} = \frac{\lambda t + s(1-\lambda t)}{1 + \lambda t - \lambda t s} = \left(\frac{\lambda t}{1+\lambda t} + s\frac{1-\lambda t}{1+\lambda t}\right)\frac{1}{1 - \frac{\lambda t}{1+\lambda t}s} \\ &= \left(\frac{\lambda t}{1+\lambda t} + s\frac{1-\lambda t}{1+\lambda t}\right)\sum_{j=0}^{\infty}\left(\frac{\lambda t}{1+\lambda t}\right)^j s^j.\end{aligned}$$

From here we have

$$\begin{aligned}p_0(t) &= \frac{\lambda t}{1+\lambda t}, \\ p_j(t) &= \frac{(\lambda t)^{j-1}}{(1+\lambda t)^{j+1}}, \quad 1 \le j < \infty.\end{aligned}$$

Now, let us consider case $\lambda \neq \mu$. The auxiliary ordinary equation is of the form

$$\frac{ds}{(\mu - \lambda s)(1-s)} = -dt$$

and solving it we obtain

$$\ln(\mu - \lambda s) - \ln(1-s) = -(\mu - \lambda)t + C,$$

the first integral is

$$\frac{\mu - \lambda s}{1-s}e^{(\mu-\lambda)t} = C_1$$

and the general solution of the partial differential equation (3.34) is

$$\Pi(s,t) = F\left(\frac{\mu - \lambda s}{1-s}e^{(\mu-\lambda)t}\right) = F\left(\frac{\mu - \lambda s}{1-s}e^{-(\lambda-\mu)t}\right).$$

From the initial condition (3.35) it follows

$$\Pi(s,0) = F\left(\frac{\mu - \lambda s}{1-s}\right) = s$$

and if we set $\frac{\mu-\lambda s}{1-s} = x$, we get

$$F(x) = \frac{\mu - x}{\lambda - x}.$$

The generating function is

$$\Pi(s,t) = \frac{\mu(1-s) - (\mu - \lambda s)e^{-(\lambda-\mu)t}}{\lambda(1-s) - (\mu-\lambda s)e^{-(\lambda-\mu)t}} = \frac{\mu(1-e^{(\lambda-\mu)t}) - s(\lambda - \mu e^{(\lambda-\mu)t})}{(\mu - \lambda e^{(\lambda-\mu)t}) - \lambda s(1-e^{(\lambda-\mu)t})}.$$

To simplify the notation let us denote

$$A(t) = \frac{1 - e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}} \ ,$$

and then after a shorter computation we can write that

$$\Pi(s,t) = \frac{1}{1-\lambda s\ A(t)}\left[\mu A(t) - \frac{\lambda - \mu e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}} s\right] = \frac{\mu A(t) - (\lambda A(t) + \mu A(t) - 1)s}{1 - \lambda s\ A(t)},$$

or

$$\Pi(s,t) = \mu A(t) + (1-\lambda A(t))(1-\mu A(t)) \sum_{j=1}^{\infty} (\lambda A(t))^{j-1} s^j.$$

From here we can conclude that

$$\begin{aligned} p_0(t) &= \mu A(t), \\ p_j(t) &= (1-\lambda A(t))(1-\mu A(t))(\lambda A(t))^{j-1}, \quad j \geq 1. \end{aligned}$$

The transition probability matrix of the embedded Markov chain $\{Y_n\}$ is

$$\boldsymbol{Q}^* = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots \\ \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & 0 & \dots \\ 0 & \frac{\mu}{\lambda+\mu} & 0 & \frac{\lambda}{\lambda+\mu} & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix},$$

we can therefore see that state 0 is absorbing; the population that will sometimes die cannot be renewed.

The probability that the population ceases to exist at time t is

$$P(X_t = 0) = p_0(t) = \begin{cases} \frac{\lambda t}{1+\lambda t}, & \lambda = \mu, \\ \frac{1-e^{(\lambda-\mu)t}}{\mu - \lambda e^{(\lambda-\mu)t}}\mu, & \lambda \neq \mu \end{cases}$$

and using limit as $t \to \infty$ we realize that the probability that the population will die out is

$$\lim_{t\to\infty} p_0(t) = \begin{cases} 1, & \lambda \le \mu, \\ \frac{\mu}{\lambda}, & \lambda > \mu. \end{cases}$$

3.8 General birth and death process

Again, let us consider a population of individuals that can reproduce and die. The size of population at time t is described by the Markov chain that is defined by intensities

$$\begin{aligned} q_{j,j+1} &= \lambda_j, \quad j = 0, 1, \dots, \\ q_{j,j-1} &= \mu_j, \quad j = 1, 2, \dots, \\ q_{jk} &= 0 \quad \text{otherwise}, \\ q_0 &= \lambda_0, \\ q_j &= \lambda_j + \mu_j, \quad j = 1, 2, \dots, \end{aligned}$$

the intensity matrix is therefore

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & 0 & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \end{pmatrix}$$

and the system of differential equations for the absolute probabilities $p_j(t)$ is of the form

$$\begin{aligned} p_0'(t) &= -\lambda_0 p_0(t) + \mu_1 p_1(t), \\ p_j'(t) &= \lambda_{j-1} p_{j-1}(t) - (\lambda_j + \mu_j) p_j(t) + \mu_{j+1} p_{j+1}(t), \quad j = 1, 2 \dots \end{aligned}$$

with the initial condition $p_i(0) = 1,\ p_j(0) = 0,\ j \ne i$.

Let us focus on the computation of limiting probabilities. We will use Theorem 3.16. We will assume that all the intensities $\lambda_j,\ j \ge 0$ and $\mu_j,\ j > 0$ are positive. Then the transition probability matrix of the embedded chain is

$$\boldsymbol{Q}^* = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots \\ \frac{\mu_1}{\lambda_1+\mu_1} & 0 & \frac{\lambda_1}{\lambda_1+\mu_1} & 0 & \dots \\ 0 & \frac{\mu_2}{\lambda_2+\mu_2} & 0 & \frac{\lambda_2}{\lambda_2+\mu_2} & \dots \\ \ddots & \ddots & \ddots & \ddots & \end{pmatrix},$$

we can thus see that the embedded chain is irreducible. To decide whether all its states are recurrent we can either use Theorem 2.23, or to find the stationary solution. If we use Theorem 2.23, we have to solve the system of equations

$$x_1 = \frac{\lambda_1}{\lambda_1 + \mu_1} x_2, \tag{3.37}$$
$$x_j = \frac{\mu_j}{\lambda_j + \mu_j} x_{j-1} + \frac{\lambda_j}{\lambda_j + \mu_j} x_{j+1},\ j \geq 2.$$

From the first equation we get

$$x_2 = \Big(1 + \frac{\mu_1}{\lambda_1}\Big) x_1$$

and further, by using the mathematical induction for $j \geq 3$

$$x_j = \Big(1 + \frac{\mu_1}{\lambda_1} + \frac{\mu_1 \mu_2}{\lambda_1 \lambda_2} + \cdots + \frac{\mu_1 \ldots \mu_{j-1}}{\lambda_1 \ldots \lambda_{j-1}}\Big) x_1.$$

If we put

$$\sigma_0 = 1, \quad \sigma_j = \frac{\mu_1 \ldots \mu_j}{\lambda_1 \ldots \lambda_j},\ j \geq 1 \tag{3.38}$$

and $x_1 = \xi$, a solution of the system (3.37) can be written in the form

$$x_1 = \sigma_0 \xi,$$
$$x_j = (\sigma_0 + \sigma_1 + \cdots + \sigma_{j-1})\xi,\ j \geq 2.$$

If $\sum_{k=0}^{\infty} \sigma_k < \infty$, then there exists a nontrivial solution of system (3.37) at the interval $[0, 1]$ (it suffices to choose $\xi = 1/\sum_{k=0}^{\infty} \sigma_k$). If $\sum_{k=0}^{\infty} \sigma_k = \infty$, then the only solution of (3.37) in interval $[0, 1]$ is the trivial one and all the states of the embedding chain are recurrent.

Thus, if the conditions of Theorem 3.16 are satisfied we can find the limit probabilities $\pi_j = \lim_{t\to\infty} p_j(t)$ by solving the system of equations $\boldsymbol{\pi}^T \boldsymbol{Q} = \boldsymbol{0}^T$, that fulfills conditions $\pi_j > 0,\ \sum_{j=0}^{\infty} \pi_j = 1$.

The considered system is

$$-\lambda_0 \pi_0 + \mu_1 \pi_1 = 0,$$
$$\lambda_{j-1}\pi_{j-1} - (\lambda_j + \mu_j)\pi_j + \mu_{j+1}\pi_{j+1} = 0, \quad j \geq 1.$$

If we put

$$\mu_j \pi_j - \lambda_{j-1}\pi_{j-1} = K_j, \quad j \geq 1,$$

we can rewrite it to

$$K_1 = 0,$$
$$K_{j+1} - K_j = 0, \quad j \geq 1.$$

Obviously, $K_j = 0$ pro $j \geq 1$, so that

$$\pi_j = \frac{\lambda_{j-1}}{\mu_j}\pi_{j-1} = \frac{\lambda_{j-1}\lambda_{j-2}}{\mu_j\mu_{j-1}}\pi_{j-2} = \cdots = \frac{\lambda_{j-1}\lambda_{j-2}\ldots\lambda_0}{\mu_j\mu_{j-1}\ldots\mu_1}\pi_0 = \rho_j\pi_0,$$

where

$$\rho_j = \frac{\lambda_{j-1}\lambda_{j-2}\ldots\lambda_0}{\mu_j\mu_{j-1}\ldots\mu_1}, \quad j \geq 1. \tag{3.39}$$

Defining $\rho_0 = 1$, we can write

$$\pi_j = \rho_j\pi_0, \quad j \geq 0.$$

The solution $\{\pi_j,\ j \geq 0\}$ is the probability distribution if and only if $\sum_{k=0}^{\infty}\rho_k < \infty$; then

$$\pi_j = \rho_j\left(\sum_{k=0}^{\infty}\rho_k\right)^{-1} \quad \text{and} \quad \pi_j > 0, \quad j \geq 0. \tag{3.40}$$

For $\lambda_0 = 0$, state 0 is absorbing and the embedded chain is reducible ($q^*_{00} = 1$); the probability that the population that at the beginning consisted exactly of i individuals sometimes will die out, is the absorbing probability at state 0 in the embedded chain starting from state i. If we denote it by u_i, we can use Theorem 2.20 and solve the system

$$u_i = q^*_{i0} + \sum_{k=1}^{\infty} q^*_{ik}u_k, \quad i = 1, 2, \ldots,$$

or

$$u_1 = \frac{\mu_1}{\lambda_1+\mu_1} + \frac{\lambda_1}{\lambda_1+\mu_1}u_2,$$
$$u_i = \frac{\mu_i}{\lambda_i+\mu_i}u_{i-1} + \frac{\lambda_i}{\lambda_i+\mu_i}u_{i+1}, \quad i \geq 2.$$

A sufficient condition the process $\{X_t\}$ to be regular is

$$\sum_{j=1}^{\infty}\frac{1}{\lambda_j} = \infty$$

(without proof).

3.9 Queueing systems

Birth and death processes are widely applicable e.g. to epidemiology or demography, here we will deal with applications to the queue theory.

By a *queueing* system we usually mean a dynamic system with a number of servers (service stations) like bank counters, cash desks, repair facilities etc. Customers are arriving into the system to be served. If the capacity of the servers is not satisfactory, the customers may wait in a queue. Being served the customers are leaving the system.

Generally we assume that the interarrival times of customers are random variables that are independent and identically distributed with a distribution A, service times of individual customers are independent and identically distributed random variables with a distribution B. The queueing system is then described by a triplet $(A/B/c)$, where c is the number of servers.

3.9.1 System $(M/M/\infty)$

The customers are arriving into the system in instants of a Poisson process with intensity $\lambda > 0$ (it means that the interarrival times are independent random variables with the exponential distribution with the mean value $\frac{1}{\lambda}$); service times are independent random variables with the same exponential distribution and the mean value $\frac{1}{\mu}, \mu > 0$. The number of service stations is so large that any customer arriving into the system is immediately starting to be served; there is no queue. If X_t denotes the number of customers in the system at time t, then $\{X_t, t \geq 0\}$ is the birth and death process with intensities

$$\begin{aligned}\lambda_j &= \lambda, \quad 0 \leq j < \infty,\\ \mu_j &= j\,\mu, \quad 1 \leq j < \infty,\end{aligned}$$

i.e., the continuous time Markov chain with the intensity matrix

$$\boldsymbol{Q} = \begin{pmatrix} -\lambda & \lambda & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & 0 & \dots \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

The differential equations for absolute probabilities $p_j(t)$ are

$$\begin{aligned}p_0'(t) &= -\lambda p_0(t) + \mu p_1(t),\\ p_j'(t) &= \lambda p_{j-1}(t) - (\lambda + j\mu)p_j(t) + (j+1)\mu p_{j+1}(t), \quad 1 \leq j < \infty\end{aligned}$$

with the initial condition $p_i(0) = 1,\ p_j(0) = 0, j \neq i$. Using the method of the generating function we obtain the partial differential equation for the generating function $\Pi(s,t)$,

$$\frac{\partial \Pi(s,t)}{\partial t} = \lambda(s-1)\Pi(s,t) - \mu(s-1)\frac{\partial \Pi(s,t)}{\partial s},$$

that is

$$\mu(1-s)\frac{\partial \Pi(s,t)}{\partial s} - \frac{\partial \Pi(s,t)}{\partial t} = \lambda(1-s)\Pi(s,t) \tag{3.41}$$

with the initial condition $\Pi(s,0) = s^i$.

A solution of this equation follows from Theorem B.8 in Appendix B. According to this theorem, we first solve the system of ordinary differential equations

$$\frac{ds}{\mu(1-s)} = -dt = \frac{d\Pi}{\lambda(1-s)\Pi};$$

solving equations

$$\frac{ds}{1-s} = -\mu dt, \qquad ds = \frac{\mu d\Pi}{\lambda \Pi}$$

we get two independent integrals

$$(1-s)e^{-\mu t} = C_1, \qquad e^{-\frac{\lambda}{\mu}s}\Pi = C_2,$$

so that the general solution of equation (3.41) is of the form

$$F[(1-s)e^{-\mu t}, e^{-\frac{\lambda}{\mu}s}\Pi] = 0$$

for a differentiable function F of two variables. From here we express Π as an implicit function of the first variable in the form

$$\Pi(s,t) = e^{\frac{\lambda}{\mu}s} f((1-s)e^{-\mu t})$$

for a differentiable function f. If we suppose that at the beginning the system was empty, (i.e., $p_0(0) = 1$), then Π fulfills the initial condition $\Pi(s,0) = 1$. Thus, function f satisfies condition

$$e^{\frac{\lambda}{\mu}s} f(1-s) = 1,$$

and with $s = 1-x$ we get

$$f(x) = e^{-\frac{\lambda}{\mu}(1-x)}.$$

Function Π is of the form

$$\Pi(s,t) = e^{-\frac{\lambda}{\mu}(1-e^{-\mu t})} e^{\frac{\lambda}{\mu}s(1-e^{-\mu t})}.$$

Expanding the second exponential function into the power series we get

$$\Pi(s,t) = e^{-\frac{\lambda}{\mu}(1-e^{-\mu t})} \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k (1-e^{-\mu t})^k s^k;$$

thus, the absolute probabilities $p_k(t)$ satisfy

$$p_k(t) = \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k (1-e^{-\mu t})^k e^{-\frac{\lambda}{\mu}(1-e^{-\mu t})}, \quad k \ \geq 0.$$

The limit probabilities are

$$\pi_k = \lim_{t\to\infty} p_k(t) = \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k e^{-\frac{\lambda}{\mu}}, \quad k \ \geq 0,$$

i.,e., the probabilities of the Poisson distribution with parameter $\frac{\lambda}{\mu}$.

The limiting probabilities can be also computed directly as follows. According to (3.38) $\sigma_0 = 1$ and for $k \geq 1$,

$$\sigma_k = \frac{\mu_1 \dots \mu_k}{\lambda_1 \dots \lambda_k} = \frac{k!\mu^k}{\lambda^k},$$

and $\sum_{k=0}^{\infty} \sigma_k = \infty$, thus, all the states of the embedded chain (that is irreducible) are recurrent. Further we proceed according to formulas (3.39) and (3.40). For $k \geq 1$ we obtain

$$\rho_k = \frac{\lambda_0 \dots \lambda_{k-1}}{\mu_1 \dots \mu_k} = \frac{\lambda^k}{k!\mu^k},$$

and setting $\rho_0 = 1$, we get

$$\sum_{k=0}^{\infty} \rho_k = \sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k = e^{\frac{\lambda}{\mu}} < \infty,$$

and from here

$$\pi_k = \frac{1}{k!} \left(\frac{\lambda}{\mu}\right)^k e^{-\frac{\lambda}{\mu}},$$

which is the previous result.

3.9.2 System $(M/M/c)$

We assume that the arrivals of customers are points of the Poisson process with intensity λ, service times are independent and identically distributed with the exponential distribution and the mean value $\frac{1}{\mu}$.

Under these assumptions, the total number X_t of customers in the service and in the queue at time t creates the continuous time Markov chain with the countable set of states $S = \{0, 1, \dots\}$ and intensities

$$q_{j,j+1} = \lambda_j = \lambda, \qquad 0 \le j < \infty,$$
$$q_{j,j-1} = \mu_j = \begin{cases} j\mu, & 1 \le j \le c, \\ c\mu, & c < j < \infty \end{cases}$$

(the other $q_{jk},\ j \neq k$ are zero). Again we have a birth and death general process with the intensities λ_j a μ_j just defined.

If we confine to the computation of limiting probabilities, formula (3.39) gives us

$$\rho_j = \begin{cases} \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j, & 1 \le j \le c, \\ \frac{c^c}{c!}\left(\frac{\lambda}{\mu c}\right)^j, & c < j < \infty; \end{cases}$$

it means that the limiting and stationary distribution exists if $\frac{\lambda}{\mu} < c$. Then

$$\pi_j = \begin{cases} \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j \frac{1}{R}, & 1 \le j \le c, \\ \frac{c^c}{c!}\left(\frac{\lambda}{\mu c}\right)^j \frac{1}{R}, & c < j < \infty \end{cases}$$

(where we have set $\rho_0 = 1$ a $R = \sum_{k=0}^{\infty} \rho_k$). The mean number of customers at the system in the stationary traffic is

$$m_1 = \sum_{j=1}^{\infty} j\pi_j \ ,$$

the mean number of customers in the queue is

$$m_2 = \sum_{j=1}^{\infty} j\pi_{j+c} \ ,$$

the mean number of the served customers is

$$m_3 = \sum_{j=1}^{c} j\pi_j + \sum_{k=1}^{\infty} c\pi_{k+c} \ ,$$

the probability that a customer will not wait for the service is

$$P_1 = \sum_{j=0}^{c-1} \pi_j,$$

and the probability of waiting for the service is $P_2 = 1 - P_1$.

Consider now the system $(M/M/c)$ with a bounded length of the queue: if there are r customers in the queue, the system is closed and other arriving customers depart without being served. The number of customers in the system is described by the Markov chain with the finite set of states and the intensity matrix

$$\begin{aligned} q_{j,j+1} &= \lambda_j = \lambda, \qquad && 0 \le j \le r+c-1, \\ q_{j,j-1} &= \mu_j = \begin{cases} j\mu, & 1 \le j \le c, \\ c\mu, & c < j \le r+c, \end{cases} \\ q_{j,k} &= 0, && j \ne k \text{ otherwise.} \end{aligned}$$

The limit distribution is $\pi_j = \rho_j / R,\ 0 \le j \le r+c$, where

$$\begin{aligned} \rho_0 &= 1, \\ \rho_j &= \begin{cases} \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j, & 1 \le j \le c, \\ \frac{c^c}{c!}\left(\frac{\lambda}{\mu c}\right)^j, & c < j \le r+c, \end{cases} \end{aligned}$$

while

$$\begin{aligned} R = \sum_{j=0}^{r+c} \rho_j &= \sum_{j=0}^{c} \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \frac{c^c}{c!}\sum_{j=c+1}^{c+r}\left(\frac{\lambda}{\mu c}\right)^j \\ &= \sum_{j=0}^{c} \frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \frac{1}{c!}\left(\frac{\lambda}{\mu}\right)^c \sum_{j=1}^{r}\left(\frac{\lambda}{\mu c}\right)^j. \end{aligned}$$

System $(M/M/1)$ with unlimited length of the queue is a special case of system $(M/M/c)$ with $c = 1$. The limit and stationary distribution does exist if $\lambda < \mu$; then

$$\rho_j = \left(\frac{\lambda}{\mu}\right)^j,\ 0 \le j < \infty, \quad R = \sum_{j=0}^{\infty} \rho_j = \left(1 - \frac{\lambda}{\mu}\right)^{-1},$$

thus

$$\pi_j = \frac{\rho_j}{R} = \left(\frac{\lambda}{\mu}\right)^j \left(1 - \frac{\lambda}{\mu}\right), \quad j \ge 0,$$

which is the geometric distribution with parameter $\frac{\lambda}{\mu}$. The mean number of customers in the system is

$$m_1 = \frac{\frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu}} = \frac{\lambda}{\mu - \lambda},$$

the mean number of customers in the queue is

$$m_2 = \sum_{k=1}^{\infty} k \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^{k+1} = \frac{(\frac{\lambda}{\mu})^2}{1 - \frac{\lambda}{\mu}} = \frac{\lambda^2}{\mu(\mu - \lambda)},$$

the mean number of customers in service is $m_3 = \frac{\lambda}{\mu}$, the probability that a customer will not wait is $P_1 = \pi_0 = 1 - \frac{\lambda}{\mu}$

Now, let us determine the mean time that a customer spends waiting in the queue and the mean time spent in the system. Assume that a customer will come at time when in the system is exactly $k \geq 1$ customers (one just served and $k-1$ waiting in the queue). (If $k = 0$, the customer will not wait and will be served immediately.)

Then the time our customer will wait in the queue is the sum of the service times of the customers standing in the queue in front of him and the remaining service time of the customer who is just served. All these random times are independent random variables with the same exponential distribution with parameter μ (we know that the exponential distribution is memoryless, thus the remaining service time has also exponential distribution). Thus the waiting time for the service of the customer standing in the queue at the $k-th$ position has the Erlang distribution of order k with the distribution function

$$F_k(x) = \begin{cases} 1 - e^{-\mu x} \sum_{j=0}^{k-1} \frac{(\mu x)^j}{j!}, & x \geq 0, \\ 0 & \text{elsewhere.} \end{cases}$$

The distribution of the waiting time T in the queue is given by the complete probability theorem:

$$\begin{aligned} F(t) = P(T \leq t) &= \pi_0 + \sum_{k=1}^{\infty} F_k(t)\pi_k = \pi_0 + \sum_{k=1}^{\infty} \left(1 - e^{-\mu t} \sum_{j=0}^{k-1} \frac{(\mu t)^j}{j!}\right) \pi_k \\ &= \pi_0 + \sum_{k=0}^{\infty} \left(1 - e^{-\mu t} \sum_{j=0}^{k} \frac{(\mu t)^j}{j!}\right) \pi_{k+1} = 1 - e^{-\mu t} \sum_{k=0}^{\infty} \sum_{j=0}^{k} \frac{(\mu t)^j}{j!} \pi_{k+1}, \end{aligned}$$

which is

$$\begin{aligned} &= 1 - e^{-\mu t} \frac{\lambda}{\mu} \left(1 - \frac{\lambda}{\mu}\right) \sum_{k=0}^{\infty} \sum_{j=0}^{k} \frac{(\mu t)^j}{j!} \left(\frac{\lambda}{\mu}\right)^k \\ &= 1 - e^{-\mu t} \frac{\lambda}{\mu} \left(1 - \frac{\lambda}{\mu}\right) \sum_{k=0}^{\infty} \frac{(\mu t)^k}{k!} \sum_{j=k}^{\infty} \left(\frac{\lambda}{\mu}\right)^j \\ &= 1 - e^{-\mu t} \frac{\lambda}{\mu} \left(1 - \frac{\lambda}{\mu}\right) \sum_{k=0}^{\infty} \frac{(\mu t)^k}{k!} \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right)^{-1} \\ &= 1 - \frac{\lambda}{\mu} e^{-(\mu - \lambda)t} = 1 - \frac{\lambda}{\mu} + \frac{\lambda}{\mu} \left(1 - e^{-(\mu - \lambda)t}\right) \end{aligned}$$

for $t \geq 0$ and it is zero for $t < 0$. From here we can easily determine that the mean waiting time in the queue is

$$w_1 = \frac{\lambda}{\mu}\frac{1}{\mu - \lambda}.$$

Notice also that the distribution function of T is not continuous at zero, $P(T = 0) = \pi_0$.

The total mean value the customer spends in the system is the mean waiting time in the queue plus the mean service time, i.e.,

$$w_2 = w_1 + \frac{1}{\mu} = \frac{1}{\mu - \lambda}.$$

Analogously we can develop the distribution of the random time the customer spends in the system: it is exponential distribution with parameter $\mu - \lambda$.

3.9.3 System $(M/G/1)$

If the distribution of the interarrival times or the distribution of the service times differs from the exponential one, the number of the customers in the system cannot be explained by a Markov chain since the Markov property is not valid, but we can still utilize some results concerning Markov chains.

System $(M/G/1)$ is a queueing system with one service station, the customers arrive according to Poisson process with intensity λ (the interarrival times have exponential distribution with the mean $\frac{1}{\lambda}$), service times are independent identically distributed random variables with a general distribution function G and a finite mean value m.

Let us consider the system at time instants when the served customers leave the system. Let X_n denote the number of customers in the system immediately after the $n-$th served customer left the system $(n \geq 1)$, let Y_n denote the number of customers that will arrive into the system during the service of the $n-$th customer. Obviously,

$$X_{n+1} = \begin{cases} X_n - 1 + Y_{n+1}, & X_n > 0, \\ Y_{n+1}, & X_n = 0. \end{cases}$$

Since the arrivals of the customers are observations of the Poisson process and we observe them in disjoint time intervals, $Y_n, n \geq 1$, are independent random variables with the same discrete distribution: for a fixed service time τ of the $n-$th customer, Y_n has the Poisson distribution with parameter $\lambda\tau$; the total probability is

$$P(Y_n = k) = \int_0^\infty e^{-\lambda\tau}\frac{1}{k!}(\lambda\tau)^k dG(\tau) = a_k,\ k \geq 0.$$

We can easily verify that $\sum_{k=0}^{\infty} a_k = 1$, hence, $\{a_k\}$ is the probability distribution. The mean value of this distribution is

$$\rho = EY_n = \sum_{k=0}^{\infty} k\ a_k = \lambda m.$$

Since random variables Y_n are independent, the sequence $\{X_n\}$ has the Markov property; we have the discrete time homogeneous Markov chain with transition probabilities

$$\begin{aligned} P(X_{n+1} = j | X_n = i) &= a_{j-i+1}, \quad i \geq 1,\ j \geq i-1, \\ &= a_j, \qquad\quad i = 0, \\ &= 0 \qquad\qquad \text{otherwise.} \end{aligned}$$

The transition probability matrix is

$$\boldsymbol{P} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

Now, let us study the problem of existence of a stationary distribution in the chain $\{X_n\}$. The system of equations $\boldsymbol{\pi}^T = \boldsymbol{\pi}^T \boldsymbol{P}$ is

$$\pi_j = \pi_0 a_j + \sum_{i=1}^{j+1} \pi_i a_{j-i+1}, \quad j \geq 0. \tag{3.42}$$

We can Solve this system of equations by the method of generating function. Denote by $A(s),\ \Pi(s)$ the generating functions of sequences $\{a_j\},\ \{\pi_j\}$, respectively. Multiplying both sides of (3.42) by s^j and add all this equations we get

$$\begin{aligned} \Pi(s) &= \pi_0 A(s) + \sum_{j=0}^{\infty} \sum_{i=1}^{j+1} \pi_i a_{j-i+1} s^j \\ &= \pi_0 A(s) + s^{-1} \sum_{i=1}^{\infty} \pi_i s^i \sum_{j=i-1}^{\infty} a_{j-i+1} s^{j-i+1} \\ &= \pi_0 A(s) + s^{-1} (\Pi(s) - \pi_0) A(s), \end{aligned}$$

from here

$$\Pi(s) = \frac{(s-1)\pi_0 A(s)}{s - A(s)}. \tag{3.43}$$

Using the limit as $s \to 1$ from the left and by using the l'Hospital rule we get

$$\lim_{s \to 1-} \Pi(s) = \sum_{j=0}^{\infty} \pi_j = \frac{\pi_0}{1-\rho}.$$

We can see that $\sum \pi_j$ is convergent if $\rho < 1$, and equals 1 for $\pi_0 = 1 - \rho$. If we insert into (3.43), we have

$$\Pi(s) = \frac{(s-1)(1-\rho)A(s)}{s - A(s)}.$$

If $\rho < 1$ the stationary distribution exists and we can obtain the probabilities π_j by expanding $\Pi(s)$. Then the mean length of the queue is $\Pi'(1)$ (left derivative).

Bibliography

[1] Billingsley, P. (1968): Convergence of Probability Measures. Wiley, New York

[2] Doob, J. (1953): Stochastic Processes. Wiley, New York

[3] Dupač, V. a Dupačová, J. (1975): Markovovy procesy I. Universita Karlova, Praha, In Czech.

[4] Dupač, V. a Dupačová, J. (1980): Markovovy procesy II. Universita Karlova, Praha, In Czech.

[5] Feller, W. (1964): An Introduction to Probability Theory and its Application Vol I. Wiley, New York.

[6] Gantmacher, F. R. (1966): Teorija matric. Nauka, Moskva. In Russian

[7] Gichman, I. I. a Skorochod, A. V. (1973): Teorija slučajnych processov II. Nauka, Moskva. In Russian

[8] Chung, K. L. (1967): Markov Chains with Stationary Transition Probabilities. Springer Verlag, New York

[9] Karlin, S. a Taylor, H. M. (1981): A Second Course in Stochastic Processes. Academic Press, New York

[10] Norris, J. R. (1997): Markov Chains. Cambridge University Press, Cambridge

[11] Prášková, Z. a Lachout P. (2012): Základy náhodných procesů I. Matfypress, Praha. In Czech.

[12] K. Rektorys a kol. (1995): Přehled užité matematiky II. Prometheus, Praha. In Czech

[13] Resnick, S. (1992): Adventures od Stochastic Processes. Birkhäuser, Boston

[14] Štěpán, J. (1987): Teorie pravděpodobnosti. Academia, Praha. In Czech.

Extending literature and applications

[15] Asmussen, S. (1987): Applied Probability and Queues. Wiley, Chichester

[16] Guttorp, P. (1995): Stochastic Modeling of Scientific Data. Chapman & Hall, London

[17] Häggström, O. (2002): Finite Markov Chains and Algorithmic Applications. Cambridge University Press, Cambridge

[18] Pardoux, E. (2008): Markov Processes and Applications. Algorithms, Networks, Genome and Finance. Wiley, Chichester

[19] Puterman, M. L. (1994): Markov Decision Processes: Discrete Stochastic Programming. Wiley, New York

[20] Rolski, T., Schmidli, H., Schmidt, V. Teugels, J. (1999): Stochastic Processes for Insurance and Finance. Wiley, Chichester

[21] Stirzaker, D. (2005): Stochastic Processes and Models. Oxford University Press, Oxford.

Appendix A

A.1 Generating functions

Definition A.1. Let $\{a_n\} = \{a_n, n \in \mathbb{N}_0\}$ be a sequence of real numbers. If the power series $A(s) = \sum_{n=0}^{\infty} a_n s^n$ converges for $|s| < s_0$ and $s_0 > 0$, we say $A(s)$ to be *generating function* of a sequence $\{a_n\}$.

Recall the basic properties of power series:

(i) To every series of type $\sum_{n=0}^{\infty} a_n s^n$ there exists a number $0 \leq R \leq \infty$, such that for $|s| < R$ this series converges absolutely; for $|s| > R$ it diverges. The number R is called radius of convergence and

$$R = \Big(\overline{\lim_{n\to\infty}}\, |a_n|^{\frac{1}{n}}\Big)^{-1}.$$

(ii) If the series $A(s) = \sum_{n=0}^{\infty} a_n s^n$ has the radius of convergence R, the series $\sum_{n=1}^{\infty} n a_n s^{n-1}$ has also the radius R and for $|s| < R$

$$\frac{d}{ds} A(s) = A'(s) = \sum_{n=1}^{\infty} n a_n s^{n-1}.$$

Then for $|s| < R$ it also holds

$$A^{(k)}(s) = \sum_{n=k}^{\infty} n(n-1)\dots(n-k+1) a_n s^{n-k}.$$

For the coefficients a_n it holds

$$a_n = A^{(n)}(0)/n!, \quad n = 0, 1, \dots$$

(iii) Abel theorem. Let the series $A(s)$ converge at the radius $R = 1$. Then the following holds true:

$$\sum_{n=0}^{\infty} a_n = a < \infty \Rightarrow \lim_{s\to 1-} A(s) = a;$$

if $a_n \geq 0 \quad \forall n \geq 0$, then

$$\lim_{s\to 1-} A(s) = \sum_{n=0}^{\infty} a_n = a \leq \infty.$$

(iv) Tauber theorem. Let the series $A(s)$ converge at the radius $R = 1$. Suppose that $a_n \geq 0 \quad \forall n \geq 0$. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} a_k = a < \infty \Leftrightarrow \lim_{s\to 1-} (1-s)A(s) = a.$$

Due to simplicity, the left limits $A(s), A'(s), A^{(k)}(s)$ as $s \to 1_-$ will be in the next denoted by $A(1), A'(1), A^{(k)}(1)$, respectively.

Let X be a nonnegative integer-valued random variable with the distribution $\{p_n, n \in \mathbb{N}_0\}$. The generating function of the sequence $\{p_n\}$ is called the *generating function of the random variable* X, we will denote it by $P(s)$, or, more precisely, $P_X(s)$. It holds $P_X(1) = \sum_{n=0}^{\infty} p_n \leq 1$, and $P_X(1) = 1$ if and only if $P[X < \infty] = 1$ (we say that X is a proper random variable). The radius of convergence of $P_X(s)$ is at least 1.

Theorem A.1. *For moments of a proper random variable X it holds*

$$EX = P'_X(1),$$

$$EX(X-1)\dots(X-k+1) = P_X^{(k)}(1),$$

$$var\, X = P''_X(1) + P'_X(1) - (P'_X(1))^2 \ (\textit{if } P'_X(1) < \infty).$$

Proof. For $0 < s < 1$ we have

$$P'_X(s) = \sum_{n=1}^{\infty} n p_n s^{n-1} = \frac{1}{s} \sum_{n=1}^{\infty} n p_n s^n = \frac{1}{s} \sum_{n=0}^{\infty} n p_n s^n$$

and from the Abel theorem applied to the sequence $\{np_n\}$ we get

$$\lim_{s\to 1-} P'_X(s) = P'_X(1) = \sum_{n=0}^{\infty} n p_n = EX.$$

Similarly,

$$\lim_{s\to 1-} P_X^{(k)}(s) = P_X^{(k)}(1) = \sum_{n=k}^{\infty} n(n-1)\dots(n-k+1)p_n = \mu^{[k]},$$

where $\mu^{[k]} = EX(X-1)\dots(X-k+1)$ is k-th factorial moment of X. □

Example A.1. The generating function of the Poisson distribution with parameter λ is

$$P_X(s) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} s^n = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda s)^n}{n!} = e^{\lambda(s-1)}.$$

From here $P_X'(s) = \lambda e^{\lambda(s-1)}$, $P_X''(s) = \lambda^2 e^{\lambda(s-1)}, \dots, P_X^{(k)} = \lambda^k e^{\lambda(s-1)}$. Thus $EX = \lambda$, $\mu^{[k]} = \lambda^k$, $\text{var } X = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Theorem A.2. *Let $P_X(1) = 1$. For $n = 0, 1, \dots,$ set*

$$q_n = P(X > n) = \sum_{j=n+1}^{\infty} p_j.$$

Let $Q(s) = \sum_{n=0}^{\infty} q_n s^n$ be the generating function of the sequence $\{q_n\}$. Then

$$Q(s) = \frac{1 - P_X(s)}{1 - s}, \quad |s| < 1. \tag{A.1}$$

The moments of the random variable X satisfy

$$EX = Q(1), \qquad \textit{var } X = 2Q'(1) + Q(1) - Q^2(1).$$

Proof. The coefficient with term s^n in the expansion $(1-s)Q(s)$ is $q_n - q_{n-1} = -p_n$, $n \geq 1$, $q_0 = 1 - p_0$. Thus

$$(1-s)Q(s) = 1 - p_0 - \sum_{n=1}^{\infty} p_n s^n = 1 - P_X(s). \tag{A.2}$$

From here we get (A.1).

Now, let us prove that $EX = Q(1)$. Relation (A.2) holds true for every $|s| < R$, where R is the radius of convergence of the series $P_X(s)$; by differentiating we get $P_X'(s) = Q(s) - (1-s)Q'(s)$. If $R > 1$, we put $s = 1$. If $R = 1$ and $EX < \infty$, we have

$$nq_n = n \sum_{k=n+1}^{\infty} p_k \leq \sum_{k=n+1}^{\infty} kp_k \to 0 \text{ as } n \to \infty$$

and $\frac{1}{n}\sum_{k=1}^{n} kq_k \to 0$. By the Tauber theorem $(1-s)Q'(s) \to 0$ as $s \to 1-$, thus $P_X'(1) = Q(1)$. If $R = 1$ and $EX = +\infty$, the assertion follows from the relation $P_X'(s) \leq Q(s)$ for $0 < s < 1$ and by using limit as $s \to 1-$. The assertion for the variance can be proved analogously. □

A.2 Convolution

Definition A.2. Let $\{a_n, n \in \mathbb{N}_0\}, \{b_n, n \in \mathbb{N}_0\}$ be sequences of real numbers. Define a sequence $\{c_n, n \in \mathbb{N}_0\}$ by

$$c_n = a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0, \quad n \geq 0.$$

The sequence $\{c_n\}$ is called to be *convolution* of the sequences $\{a_n\}, \{b_n\}$, we denote $\{c_n\} = \{a_n\} * \{b_n\}$.

Theorem A.3. *Let $\{a_n\}, \{b_n\}$ be real-valued sequences with the generating functions A, B, respectively. Then the generating function C of their convolution $\{c_n\}$ is*

$$C(s) = A(s)B(s).$$

Proof. It follows by the multiplication of the power series $A(s)$ and $B(s)$ and the comparison the coefficients with the same powers of s. □

It can be easily shown that the convolution operation is associative and commutative. The sequence $\{a_n\} * \{a_n\}$ is called *the second convolutional power* of the sequence $\{a_n\}$, denoted by $\{a_n\}^{2*}$. Similarly we define the *k-th convolutional power* by $\{a_n\}^{k*} = \{a_n\}^{(k-1)*} * \{a_n\}$. Let us define $\{a_n\}^{0*} = \{1, 0, 0, \dots\}$; the $\{a_n\}^{1*} = \{a_n\}$. If $A(s)$ is the generating function of the sequence $\{a_n\}$, then $A^k(s)$ is the generating function of the sequence $\{a_n\}^{k*}$.

Let X_1, X_2 be independent integer-valued random variables with respective distributions $\{p_n^{(1)}\}, \{p_n^{(2)}\}$. Let $S = X_1 + X_2$. Then the distribution of random variable S is $\{p_k\}$, where

$$p_k = P(S = k) = \sum_{j=0}^{k} P(X_1 = j)P(X_2 = k - j).$$

Thus, $\{p_n\} = \{p_n^{(1)}\} * \{p_n^{(2)}\}$ and the generating function of S is

$$P_S(s) = P_{X_1}(s)P_{X_2}(s).$$

Especially, if X_1, X_2 are independent and identically distributed with the distribution $\{a_n\}$, the distribution of their sum is the second convolutional power of the sequence $\{a_n\}$. This assertion can be generalized to any finite sum of independent identically distributed random variables.

Definition A.3. Let F be a distribution function of a nonnegative random variable, i.e., $F(x) = 0, x < 0$, and let G be a function defined on $\mathbb{R}_+ = [0, \infty)$, that is locally bounded, i.e., bounded on every finite interval. *Convolution of functions F, G* is defined to be

$$(F * G)(t) = \int_0^t G(t - x)dF(x), \quad t \geq 0.$$

Theorem A.4. *$F * G \geq 0$, if G is nonnegative, and $F * G$ is locally bounded..*

Proof. It is obvious that $F * G$ is nonnegative. Further, for every $0 \leq s \leq t$

$$|(F * G)(s)| \leq \int_0^t |G(s - x)dF(x)| \leq \sup_{0 \leq s \leq t} |g(s)| \int_0^s dF(s) \leq \sup_{0 \leq s \leq t} |g(s)|F(t) < \infty.$$

□

Function $F * F$ is called *second convolutional power of function F,*, denoted F^{2*}. Define a general convolutional power by

$$\begin{aligned} F^{0*}(x) &= I([0, \infty))(x), \\ F^{1*}(x) &= F(x), \\ F^{(n+1)*}(x) &= (F^{n*} * F)(x), \quad n \geq 1. \end{aligned}$$

Theorem A.5. *Let X_1, X_2 be independent random variables taking only nonnegative values, X_1 with a distribution function F_1, X_2 with a distribution function F_2, respectively. Then the distribution function of $X_1 + X_2$ is $F_1 * F_2$.*

Proof. For $t \geq 0$

$$P(X_1 + X_2 \leq t) = \int\int_{0 \leq x+y \leq t} dF_1(x)dF_2(y)$$
$$= \int_0^t \left(\int_0^{t-x} dF_2(y) \right) dF_1(x) = \int_0^t F_2(t-x)dF_1(x).$$

□

From the proof of A.5 we can see that $F_1 * F_2 = F_2 * F_1$. If X_1, X_2 are equally distributed with the distribution F, then $X_1 + X_2$ has distribution function F^{2*}. Generalization to any finite sum of independent identically distributed random variables is obvious.

A.3 Laplace transformation

Let X be a nonnegative random variable with a distribution function F. Then the Laplace transformation of the distribution function F is

$$\widehat{F}(\lambda) = \int_0^\infty e^{-\lambda x} dF(x) = Ee^{-\lambda X}, \quad \lambda \geq 0.$$

Obviously, $\widehat{F}(\lambda) < \infty$ for all $\lambda \geq 0$.

Theorem A.6. *Let X_1, X_2 be independent random variables with distribution functions F_1, F_2, respectively. Then*

$$(\widehat{F_1 * F_2})(\lambda) = \widehat{F}_1(\lambda)\hat{F}_2(\lambda).$$

If X_1, X_2 are equally distributed then

$$(\widehat{F * F})(\lambda) = \widehat{F^{*2}}(\lambda) = \left(\hat{F}(\lambda)\right)^2.$$

Proof. The assertion follows from the independence of random variables and properties of the mean value.

□

A.4 Random sum of random variables

Let $N, X_1, X_2, \dots$ be independent integer-valued random variables, N with the distribution $\{\pi_n, n \in \mathbb{N}_0\}$ and all X_i have the same distribution $\{p_n, n \in \mathbb{N}_0\}$. Set $S_0 = 0, S_N = X_1 + \cdots + X_N, N \geq 1$. Denote the distribution of S_N by $\{h_n, n \in \mathbb{N}_0\}$.

Theorem A.7. *Suppose that the generating function of random variable N is Π, and the generating function of random variables X_i is P. Let H be the generating function of random variable S_N. Then $H(s) = \Pi(P(s))$ and*

$$ES_N = ENEX_1,$$

$$var\ S_N = (EN)(var\ X_1) + (\ var\ N)(EX_1)^2.$$

Proof. By the complete probability theorem and due to independence

$$\begin{aligned} h_k := P(S_N = k) &= \sum_{n=0}^{\infty} P(S_N = k|N = n)P(N = n) \\ &= \sum_{n=0}^{\infty} P(S_n = k|N = n)P(N = n) \\ &= \sum_{n=0}^{\infty} P(S_n = k)P(N = n) = \sum_{n=0}^{\infty} p_k^{n*}\pi_n, \end{aligned}$$

where p_k^{n*} is k-th element of the sequence $\{p_k\}^{n*}$. Then

$$H(s) = \sum_{k=0}^{\infty} h_k s^k = \sum_{k=0}^{\infty}\sum_{n=0}^{\infty} p_k^{n*}\pi_n s^k = \sum_{n=0}^{\infty} \pi_n P^n(s) = \Pi(P(s)).$$

For the mean value of S_N we have

$$ES_N = H'(1) = \Pi'(P(1))P'(1) = \Pi'(1)P'(1) = ENEX_1,$$

and the formula for the variance follows analogously from Theorem A.1.

□

Appendix B

B.1 Some results from matrix theory

Definition B.1. Let $\sum_{n=0}^{\infty} a_n t^n$ be power series in $\mathbb{C}$ and let $\boldsymbol{A}$ be a squared matrix of order k. Then the series

$$\sum_{n=0}^{\infty} a_n \boldsymbol{A}^n = a_0 \boldsymbol{I} + a_1 \boldsymbol{A} + a_2 \boldsymbol{A}^2 + \dots$$

is called to be *power series* of matrix $\boldsymbol{A}$. If the sequence of partial sums

$$\boldsymbol{S}_N = \sum_{n=0}^{N} a_n \boldsymbol{A}^n$$

converges to a matrix $\boldsymbol{S}$ element-wise, we say that the series $\sum_{n=0}^{\infty} a_n \boldsymbol{A}^n$ converges and its sum is $\boldsymbol{S}$.

Theorem B.1. *If the power series $\sum_{n=0}^{\infty} a_n t^n$ has convergence radius $R > 0$ and all the eigenvalues of the matrix $\boldsymbol{A}$ are in modulus less then R, then the power series $\sum_{n=0}^{\infty} a_n \boldsymbol{A}^n$ converges.*

Proof. Gantmacher (1966), Chap. V, § 4, p. 117. □

Theorem B.2. *Let $\boldsymbol{A}$ be a squared matrix such that $\boldsymbol{A}^n \to \boldsymbol{0}$ as $n \to \infty$, where $\boldsymbol{0}$ is the null matrix. Then the matrix $\boldsymbol{I} - \boldsymbol{A}$ is regular and*

$$(\boldsymbol{I} - \boldsymbol{A})^{-1} = \sum_{k=0}^{\infty} \boldsymbol{A}^k. \tag{B.1}$$

Proof. For every $n \geq 1$ it holds

$$\boldsymbol{I} - \boldsymbol{A}^n = (\boldsymbol{I} - \boldsymbol{A})(\boldsymbol{I} + \boldsymbol{A} + \cdots + \boldsymbol{A}^{n-1}).$$

Further, from the assumption of the theorem it follows $\boldsymbol{I} - \boldsymbol{A}^n \to \boldsymbol{I}$ as $n \to \infty$, and since the determinant is a continuous function of the elements of the matrix, it also holds $|\boldsymbol{I} - \boldsymbol{A}^n| \to |\boldsymbol{I}|$. Thus, there exists $n_0 in \mathbb{N}$ such that $|\boldsymbol{I} - \boldsymbol{A}^{n_0}| \neq 0$. Then

$$|\boldsymbol{I} - \boldsymbol{A}^{n_0}| = |\boldsymbol{I} - \boldsymbol{A}|.|\boldsymbol{I} + \boldsymbol{A} + \cdots + \boldsymbol{A}^{n_0-1}| \neq 0,$$

which implies that the matrix $\boldsymbol{I} - \boldsymbol{A}$ is regular and its inverse matrix does exist. Thus, for every $n \geq 1$

$$(\boldsymbol{I} - \boldsymbol{A})^{-1}(\boldsymbol{I} - \boldsymbol{A}^n) = \boldsymbol{I} + \boldsymbol{A} + \cdots + \boldsymbol{A}^{n-1},$$

and using limit we get

$$(\boldsymbol{I} - \boldsymbol{A})^{-1} = \lim_{n\to\infty} \sum_{k=0}^{n-1} \boldsymbol{A}^k = \sum_{k=0}^{\infty} \boldsymbol{A}^k.$$

□

Definition B.2. We say that a squared matrix $\boldsymbol{A}$ is *reducible,* if (after a permutation of the rows and columns) it can be written in the form

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{A}_2 & \boldsymbol{A}_3 \end{pmatrix},$$

where $\boldsymbol{0}$ is the null matrix and $\boldsymbol{A}_1$ and $\boldsymbol{A}_3$ are squared matrices. In the opposite case, $\boldsymbol{A}$ is *irreducible.*

Theorem B.3. Perron–Frobenius *Every irreducible matrix with nonnegative elements has a positive eigenvalue r, that is simple, and all the elements of the respective eigenvector are positive. Other eigenvalues are at most equal to r in the modulus.*

If $\boldsymbol{A}$ is an irreducible squared stochastic matrix, all the eigenvalues of $\boldsymbol{A}$ are less then 1 in modulus. Number 1 *is the simple eigenvalue of $\boldsymbol{A}$, and all the elements of the respective eigenvector are equal to the same constant.*

Proof. Gantmacher (1966), Chap. XIII, § 2, pp. 354-355 a 382. □

Theorem B.4. *Let $\boldsymbol{A}$ be squared matrix with elements a_{ij} such that $a_{ii} \leq 0, a_{ij} \geq 0$, $i \neq j$ and all row sums are equal to* 0*. Then* 0 *is the eigenvalue of the matrix $\boldsymbol{A}$ with the eigenvector* $(1,\ldots,1)^T$*, and the other eigenvalues λ of $\boldsymbol{A}$ satisfy* $Re\,\lambda < 0$.

Proof. Dupač, Dupačová II (1980), p. 43. □

Definition B.3. For a squared matrix $\boldsymbol{A} = \{a_{ij}, 1 \leq i, j \leq k\}$ we define *adjoint matrix* $\mathrm{adj}\,(\boldsymbol{A}) = \{b_{ij}, 1 \leq i, j \leq k\}$ by

$$b_{ij} = (-1)^{i+j} \det\{a_{rs}, 1 \leq r, s \leq k, r \neq j, s \neq i\}. \tag{B.2}$$

Theorem B.5. *For a squared matrix $\boldsymbol{A}$,*

$$\boldsymbol{A}\, adj\,(\boldsymbol{A}) = adj\,(\boldsymbol{A})\,\boldsymbol{A} = (\det \boldsymbol{A})\,\boldsymbol{I},$$

where $\boldsymbol{I}$ is the identity matrix.

Proof. It follows from the properties of the inverse matrix. □

Definition B.4. Let $0 < R \leq \infty$. For a holomorphic function $f : \mathcal{U}(0,R) \to \mathbb{C}$ on a neighbourhood $\mathcal{U}(0,R)$, i.e., having on $\mathcal{U}(0,R)$ derivatives of all orders, we define an extension to all squared matrices $\boldsymbol{A}$, the eigenvalues of them are less than R in modulus, by

$$f(\boldsymbol{A}) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} \boldsymbol{A}^k\ .$$

Theorem B.6. Perron formula *Let $f : \mathcal{U}(0,R) \to \mathbb{C}$ be a holomorphic function on the neighbourhood $\mathcal{U}(0,R)$ for some $0 < R \leq \infty$. Let $\boldsymbol{A}$ be a squared matrix, the eigenvalues of which are $\lambda_1,\ldots,\lambda_k$ with multiplicities $m_1,\ldots,m_k$, respectively. If $|\lambda_j| < R$ for all $j = 1,2,\ldots,k$, then*

$$f(\boldsymbol{A}) = \sum_{j=1}^{k} \frac{1}{(m_j-1)!} \frac{d^{m_j-1}}{d\lambda^{m_j-1}} \left[\frac{(\lambda-\lambda_j)^{m_j}}{\det(\lambda \boldsymbol{I} - \boldsymbol{A})} f(\lambda)\ adj\ (\lambda \boldsymbol{I} - \boldsymbol{A}) \right] \Bigg|_{\lambda=\lambda_j}\ . \tag{B.3}$$

Proof. Gantmacher (1966), Chap. V, § 3, p. 113. □

Special cases of the Perron formula:

$$\boldsymbol{A}^n = \sum_{j=1}^{k} \frac{1}{(m_j-1)!} \frac{d^{m_j-1}}{d\lambda^{m_j-1}} \left[\frac{\text{adj } (\lambda \boldsymbol{I} - \boldsymbol{A})}{\psi_j(\lambda)} \lambda^n \right] \Bigg|_{\lambda=\lambda_j} .$$

$$e^{\boldsymbol{A}} = \sum_{j=1}^{k} \frac{1}{(m_j-1)!} \frac{d^{m_j-1}}{d\lambda^{m_j-1}} \left[\frac{\text{adj } (\lambda \boldsymbol{I} - \boldsymbol{A})}{\psi_j(\lambda)} e^{\lambda} \right] \Bigg|_{\lambda=\lambda_j} ,$$

where we denoted

$$\psi_j(\lambda) = \frac{\det(\lambda \boldsymbol{I} - \boldsymbol{A})}{(\lambda - \lambda_j)^{m_j}}.$$

If all the eigenvalues of the matrix $\boldsymbol{A}$ are simple, formula (B.3) can be written in a simplified form

$$f(\boldsymbol{A}) = \sum_{j=1}^{K} \frac{\text{adj } (\lambda_j \boldsymbol{I} - \boldsymbol{A})}{\psi_j(\lambda_j)} f(\lambda_j) \ , \tag{B.4}$$

where K is the order of the matrix.

B.2 Partial differential equations

Definition B.5. An ordinary differential equation of order n is the equation

$$F(x, y, y', \dots, y^{(n)}) = 0, \tag{B.5}$$

or, if it is solved with respect to the highest order derivative, the equation

$$y^{(n)} = f(x, y, y', \dots, y^{(n-1)}). \tag{B.6}$$

A solution (integral) of this equation is any function $y = g(x)$, that has derivatives up to order n and satisfies identically equation (B.5). The solution is often given implicitly in the form $\phi(x, y) = c$, where c is a constant.

Definition B.6. The partial differential equation is a relation between an unknown function $z(x_1, \dots, x_n)$ of variables $x_1, \dots, x_n, n \geq 2$ and its derivatives given by

$$F\left(x_1, \dots, x_n, z, \frac{\partial z}{\partial x_1}, \dots, \frac{\partial z}{\partial x_n}, \frac{\partial^2 z}{\partial x_1^2}, \dots, \frac{\partial^k z}{\partial x_1^k}, \dots\right) = 0. \tag{B.7}$$

The order of this equation is the order of the highest derivative that appears in this equation.

A solution (integral) of the equation is any differentiable function $z(x_1, \dots, x_n)$, that satisfies (B.7) at every point $x_1, \dots, x_n$.

Definition B.7. A homogeneous linear partial differential equation of the first order in two variables x, y is the equation

$$a(x,y)\frac{\partial z}{\partial x} + b(x,y)\frac{\partial z}{\partial y} = 0, \tag{B.8}$$

where a, b are continuous functions in the respective region and are not simultaneously equal to zero in it.

Theorem B.7. *Let $\psi(x,y) = c$ be the first integral of the equation*

$$\frac{dx}{a(x,y)} = \frac{dy}{b(x,y)}.$$

The the solution to equation (B.8) is

$$z = F(\psi(x,y)),$$

where F is an arbitrary differentiable function.

(Rektorys, 1995, Vol II, Sect. 18.2, Theorem 1 (without proof).)

Definition B.8. A nonhomogeneous linear partial differential equation of the first order in two variables x, y is the equation

$$P(x,y,z)\frac{\partial z}{\partial x} + Q(x,y,z)\frac{\partial z}{\partial y} = R(x,y,z). \tag{B.9}$$

Functions P, Q, R are supposed to be continuous in the given region, P, Q are not simultaneously equal to zero at this region and $R \not\equiv 0$.

Theorem B.8. *Let $\phi(x,y,z) = c_1$ and $\psi(x,y,z) = c_2$ be two independent integrals of the system of equations*

$$\frac{dx}{P(x,y,z)} = \frac{dy}{Q(x,y,z)} = \frac{dz}{R(x,y,z)}.$$

Then the solution of equation (B.9) is given in an implicit form by

$$F(\phi(x,y,z), \psi(x,y,z)) = 0,$$

where F is an arbitrary differentiable function of two variables.

(Rektorys, 1995, Vol II, Sect. 18.2, Theorem 2 (without proof).)