

Spatial Queues

Lothar Breuer
University of Trier

Contents

1. Introduction	1
1.1. Summary	2
1.2. Notations and Conventions	4
1.3. Acknowledgements	5
2. Markov-Additive Jump Processes	7
2.1. Definition	8
2.2. Elementary Properties	12
2.3. MAJPs on a Real Vector Space	14
2.4. Markovian Arrival Processes	19
2.5. Laws of Large Numbers	21
3. Spatial Markovian Arrival Processes	25
3.1. Definition and Construction	25
3.2. Properties	28
3.2.1. Marginal Processes	28
3.2.2. Transition Probabilities	29
3.2.3. Z-Transform and Expectation	32
3.2.4. Periodic SMAPs	33
3.3. Examples	35

4. Markovian Spatial Queues	39
4.1. The $SMAP/M_t/c/c$ Queue	40
4.1.1. Transient Distribution	41
4.1.2. Asymptotic Distributions	44
4.1.3. A Loss Formula	46
4.2. The Periodic $BMAP/M_t/c$ Queue	49
4.2.1. Transient Distributions	49
4.2.2. Stability and Asymptotic Distributions	50
5. Spatial Queues with Infinitely Many Servers	55
5.1. Homogeneous Arrival Rates without User Movements	55
5.1.1. IID Service Time Distributions	56
5.1.2. Spatially Variable Service Time Distributions	60
5.1.3. Common Distribution in Finitely Many Subsets	60
5.1.4. Stability and Asymptotic Distribution	62
5.2. General Arrival Rates without User Movements	66
5.2.1. Transient Distribution	66
5.2.2. An Approximation	67
5.3. General Arrival Rates with User Movements	70
5.4. Application: Planning a Mobile Communication Network	72
6. Parameter Estimation	75
6.1. Arrival Rates	76
6.2. Phases at Arrival Instants	77
6.3. Generator Matrix	78
6.4. Estimation for BMAPs	79
6.5. Remarks	81

Appendix	85
A. Markov Jump Processes	87
A.1. Definition and Properties	87
A.2. Quasi-Commutability of Generators	103
A.3. Periodic Markov Jump Processes	108
B. Stochastic Point Fields	115
B.1. Set Version	115
B.2. Version for Counting Functions	116
B.3. Construction	118
B.4. Examples	120

Contents

1. Introduction

The common gains from communication between mathematical theory and practical application can be traced in the field of queueing theory especially well. Starting with Erlang's research at the beginning of the 20th century, this whole research area was born as the endeavour of building appropriate stochastic models for the performance analysis of technical systems. The initial application area of telephone systems was soon enlarged by applications in machine repair, inventory control, assessment of insurance risk and later the design and analysis of computer systems, to name but a few. The close interaction between theory and practice has remained a driving force for the development of queueing theory until today.

In this respect, the momentum coming from applicational demands in the field of telephone networks has not lost its strength. The integrated transmission of voice, data and video sources via the same channel crucially contributed to the shift from telephone networks of Erlang's times towards telecommunication systems of today. This required the use of more sophisticated models for data streams than the now already classical Poisson process could provide. At the end of the 1970s, Neuts [71] proposed a richer class of processes, which under the name "Batch Markovian Arrival Processes" (shortly BMAPs) was notationally simplified by Lucantoni [60] in 1991. This class of processes, which actually is a subset of the class of Markov-additive processes introduced by Çinlar [25] and even earlier by Ezhov, Skorokhod [37] or Neveu [73], engendered a new direction in queueing theory using so-called matrix-analytical methods.

The present thesis is motivated by a further step in the development of modern telecommunication systems. The technical availability of wireless data transmission opened up the way for mobile communication systems, which today are used by increasingly many subscribers. At the same time, this development exhibited the need for queueing models that can reflect the spatial distribution of the users in the system. These are indispensable in order to model phenomena like spatially dependent service time distributions, user movements or the spatial distribution of users in the system. The analysis of spatial queues yields such essential values as the necessary cell capacities needed to guarantee a low outage probability or the outage probability of an existing network. Such knowledge greatly improves the competence of planning and operating mobile communication networks.

There have been several authors who introduced queueing models designed to explore the spatial properties of mobile communication systems. Some of the recent notable contributions are the following:

1. Introduction

In 1995, Çinlar [30] used the concept of Poisson random measures in order to obtain distributions for the spatial $M/G/\infty$ queue with Poisson arrivals in time and space. Furthermore, mobility of customers is modelled using a method from Massey, Whitt [63], namely integrating the stochastic location process with respect to the Poisson arrival process. The same approach will be used in this thesis for a more general model (see section 5.3).

Two years later, Baccelli, Zuyev [9] introduced a sophisticated model for mobile communication networks using mainly concepts from the theory of point processes. Analytical results are given only for models which are based on the Poisson process. One of the objectives of the present thesis is to obtain explicit results for models more general than Poisson models.

Boucherie, Dijk, Mandjes [21, 22] developed a model for cellular mobile communication networks which is based on Jackson networks and more general product form networks. Thus the analysis can be performed using known methods for queueing networks. The spatial distribution has a granularity as fine as the number of cells in the network.

The crucial ideas for large parts of this thesis came from Baum [16, 17]. In these papers, a concept of spatial BMAPs is developed and the respective infinite server queue is analysed. It is shown that this spatial generalization of BMAPs retains the properties of BMAPs which yield a tractable analysis of such arrival processes and their respective queues. These ideas are taken up in this thesis and lead to the construction of the so-called Spatial Markovian Arrival Process (SMAP). It turns out that the analytical methods developed in Baum [14, 16, 17], which were intended to be applied to queues with BMAP input, can be adapted to the analysis of several practically relevant queues with SMAP input.

The aim of the present thesis in this context is to develop a unified theory of spatial queues, which is broad enough to serve as a base for queueing models arising from applicational needs in mobile communication networks, but also specific enough to yield concrete results for the performance analysis of such networks. A particular objective is to develop the most natural generalization of existing concepts (e.g. the BMAP) toward the needs of mobile communication networks. To these belong the spatial distribution of batch arrivals and users in the system as well as time-inhomogeneous (e.g. periodic) arrival intensities and user movements. The concepts used in this thesis will be embedded into existing mathematical concepts. Thus it is possible to show the similarity of those to concepts used in classical queueing theory. Furthermore, classical mathematical results can be applied in order to analyze spatial queues.

1.1. Summary

In this thesis, the groundwork is laid for designing queueing models for the application field of mobile communication networks. Included are the following new concepts and results: First, a class of spatial arrival processes (called Spatial Markovian Arrival Processes or shortly SMAPs) is defined, which is general enough to model spatially distributed batch arrivals as well as time-inhomogeneous (e.g. periodic) arrival rates and a general phase space. It is shown that SMAPs are a natural generalization of BMAPs and retain the typical properties of

BMAPs, which make them suitable for the matrix convolution calculus introduced by Baum [14, 15].

Then some practically relevant queues, which are fed by SMAP inputs, are analyzed. These queues feature the definition of spatially variable service time distributions, i.e. the service time of a user may depend on its location at the time instant of arrival. Furthermore, their analysis yields the spatial distribution of users in the system in the transient as well as in the stationary case. Where applicable, loss formulae are given. A further new feature of the queueing models presented in this thesis is the possibility of including time-inhomogeneous arrival intensities (e.g. periodic ones) and deriving the resulting transient and asymptotic distributions for the respective queues. Since in mobile communication networks there are no waiting users (if the line is busy, a user would give up or try again later), only spatial queues with either infinitely many servers or zero waiting room capacity are examined.

In order to practically use stochastic models, it is necessary to obtain statistically validated estimates of the model parameters. Hence, a routine for parameter estimation from empirical arrival streams is derived in the present thesis, too. This estimation procedure is applicable also for BMAPs, for which it is the only published estimation procedure up to now.

The main theoretical concepts, which are used in this thesis, are the theory of Markov jump processes and the concept of stochastic point fields and their representation by counting functions. Both are shortly introduced in appendices A and B. For special jump processes, new results are presented in sections A.2 and A.3.

In mobile communication networks the integrated transmission of different data streams is of standard use. Therefore, it seems appropriate to base models of such networks on the BMAP concept in order to preserve the capability of modelling integrated service transmission. In the present thesis, before application-oriented models are developed in terms of spatial queues, the concept of BMAPs is generalized towards a special class of Markov jump processes in chapter 2. For this class of so-called Markov-additive jump processes, several properties are derived which allow the same methods of analysis to be used for spatial queues (as set forth in this thesis) as well as for the very general class of queues on real vector spaces (instead of the classical queues on the space of non-negative integers). This opens the way for the application of queueing theory to many fields which cannot be modelled by classical one-dimensional queues. Besides the extension of applicability of the theory, this chapter intends to exhibit fundamental results on arrival processes on a fairly abstract level in order to further explore the nature of Markov-additive jump processes as an own interest.

In chapter 3, the class of Spatial Markovian Arrival Processes (SMAPs) is defined. This class will provide the arrival processes for all queues analyzed in this thesis. SMAPs are a generalization of BMAPs which creates the possibility of modelling spatially distributed batch arrivals, time-inhomogeneous (e.g. periodic) arrival rates and a general phase space. Although this class of arrival processes is much more general than the class of BMAPs, it retains the same crucial properties that make their analysis tractable. This is shown by the derivation of the most important properties of SMAPs. At the end of this chapter, some examples for SMAPs are given, among them the class of non-homogeneous (especially periodic) BMAPs.

1. Introduction

Queues with SMAP arrival processes will be treated as spatial queues. The class of Markovian spatial queues (i.e. those queues which are Markov processes) is treated in chapter 4. For ease of notational simplicity, this case is examined not in the most general setting, but in terms of the $SMAP/M_t/c/c$ queue. Yet the same method of analysis applies to the $SMAP/M_t/c/c$ queue with spatially variable service time distribution or the $SMAP/PH/c/c$ queue in a completely analogous way. Since these queues are loss systems, a formula for the loss probability is derived, too. As an example for the utility of the concept of non-homogeneous BMAPs, the periodic $BMAP/M_t/c$ queue is examined by the same analytical methods. Like all the spatial queues introduced in the present thesis, this queue has not been analyzed before.

Chapter 5 covers the class of spatial infinite server queues. The method of analysis applied to this class allows a variety of features of mobile communication networks to be included in the model. The service time distribution may be general and spatially variable. Furthermore, the arrival rates may depend on time (e.g. periodic) and even stochastic user movements can be modeled in a natural way. A simple formula for the asymptotic distribution can be given only for the case of homogeneous arrival rates without user movements. For the other cases, approximations along with an estimation of the approximation error are given.

In the last chapter of this thesis, a routine for parameter estimation is introduced for SMAPs. Although crucial for purposes of practical application, the problem of finding some statistically validated parameter estimation has not been solved yet, even for BMAPs. The routine introduced here is divided into three steps, each of which uses classical statistical methods of inference. It can further be applied to BMAPs for which no other estimation routine has been proposed yet.

1.2. Notations and Conventions

The set of positive integers is denoted by \mathbb{N} and the set of non-negative integers by \mathbb{N}_0 . The positive real numbers are denoted by \mathbb{R}^+ and the non-negative real numbers by \mathbb{R}_0^+ . The empty sum as well as the empty product shall always be defined as the neutral element. Random variables are defined on an underlying probability space which is not mentioned explicitly. The end of a proof is marked by a smiley.

Let (S, \mathcal{S}) denote a measurable space and Q be a kernel on $S \times S$. The k -th iteration Q^k of Q shall be defined by $Q^0 := I$ and iteratively

$$Q^{k+1}(x, A) := \int_S Q^k(x, dy) Q(y, A)$$

for all $x \in S$ and $A \in \mathcal{S}$, with I denoting the identity kernel. The latter is defined as

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

for all $x \in S$ and $A \in \mathcal{S}$.

In the case of a discrete state space, the states will be denoted by letters as $h, i, j, k \in S$ and we write $Q(i, j)$ instead of $Q(i, \{j\})$.

1.3. Acknowledgements

The fundamental idea of how to construct Spatial Markovian Arrival Processes goes back to Prof. Dieter Baum, who thus inspired the topic of the present thesis. Through many helpful discussions, he continued to motivate and support this work during its process of development. For this, I would like to express my gratitude and appreciation. Further, I would like to thank Prof. Igor Kovalenko for refereeing this thesis and for his helpful comments, which led to substantial improvements in the final version of this work. Many thanks go to my colleague Thomas Perst for making the smiley at the end of the proofs and other helpful hints of how to handle the \LaTeX system. Last but not least, I would like to appreciate the patience of my friend Corinna, who had to suffer many times of absentmindedness and solitary mood, which seem to be characteristic byproducts of mathematical research.

1. Introduction

2. Markov-Additive Jump Processes

Inspired by a process developed in Ezhov, Skorokhod [37], the term "Markov-additive process" has been coined by Çinlar [25] in 1972. It refers to a two-dimensional Markov process with transition probabilities that depend on one dimension only. The marginal process in this dimension is a Markov process, too, and shall be called phase process in the present thesis. The marginal process in the other dimension is a process with conditionally independent increments given the phase process.

The first idea of such a process goes back to Neveu [73] in 1961, whose so-called F-process is the class of Markov-additive processes with a finite phase space. A special case of F-processes is the class of BMAPs developed in Lucantoni [60] thirty years later, which are now widely used in queueing theory. The founding work of Çinlar [25] on Markov-additive processes was followed by several studies in the 1970s and 80s, e.g. Arjas, Speed [3], Çinlar [27], [28], [29], or Ney, Nummelin [75] for discrete time. This chapter mainly goes along with the paper by Pacheco, Prabhu [77], which was written after the BMAP concept proved successful as a versatile arrival process for queues. The concept developed in Pacheco, Prabhu [77] will be slightly generalized in this chapter and, most importantly, the inhomogeneous case will be analyzed, too.

In this chapter, the focus is on Markov-additive processes which belong to the class of Markov jump processes, since they are sufficient as arrival processes for queueing theory. In the first section, these so-called Markov-additive jump processes are defined and the transition probabilities are derived in terms of the infinitesimal transition rates. The next section states some elementary properties, mostly resulting from the definition immediately. Under the assumption that the additive part of the state space is a real vector space, Fourier transforms and expectations are derived in section 2.3. The more specific assumption that the increments for the additive part be non-negative integers in every component of the vector space leads to the same derivations as in section 2.3 via the easier notion of z-transforms. These processes will be called Markovian arrival processes and analyzed in section 2.4. In the last section of this chapter, laws of large numbers are given for the special cases of periodic or homogeneous Markov-additive jump processes on real vector spaces.

2.1. Definition

Markov-additive jump processes will be defined as two-dimensional Markov jump processes which satisfy the condition that the transition probabilities depend on the second dimension only. The first dimension is an (additive) semi-group and the marginal process on it turns out to have conditionally independent increments given the marginal process in the second dimension.

Definition 2.1 Let $(N, J) = ((N_t, J_t) : t \in \mathbb{R}_0^+)$ be a two-dimensional Markov jump process with state space $S := \Sigma \times \Phi$. Let $\sigma(\Sigma)$ and $\sigma(\Phi)$ be σ -algebras on Σ and Φ , respectively, which satisfy $\{x\} \in \sigma(\Sigma)$ and $\{y\} \in \sigma(\Phi)$ for all $x \in \Sigma$ and $y \in \Phi$. Denote $\mathcal{S} := \sigma(\Sigma) \otimes \sigma(\Phi)$ as the product σ -algebra of $\sigma(\Sigma)$ and $\sigma(\Phi)$. Further, let $(\Sigma, +)$ be a semi-group with neutral element $0 \in \Sigma$. Assume that for all $x \in \Sigma$ and $A \in \sigma(\Sigma)$, we also have $A - x := \{s \in \Sigma : s + x \in A\} \in \sigma(\Sigma)$. Then (N, J) is called **Markov-additive jump process** if the transition probabilities satisfy the condition

$$P(s, t; (x, y), A \times B) = P(s, t; (0, y), (A - x) \times B) \quad (2.1)$$

for all $s < t \in \mathbb{R}_0^+$, $(x, y) \in S$ and $A \times B \in \mathcal{S}$. Define

$$P(s, t; y, A \times B) := P(s, t; (0, y), A \times B)$$

for all $s < t \in \mathbb{R}_0^+$, $y \in \Phi$ and $A \times B \in \mathcal{S}$.

Remark 2.1 Definition A.1 postulates that the state space $S := \Sigma \times \Phi$ of a Markov-additive jump process be locally compact, separable and metric. Standard results in topology (cf. Herrlich [48], p.224,118,117) yield that S satisfies this condition if and only if the state spaces Σ and Φ of the marginal processes do so.

Remark 2.2 Because of equation 2.1, a Markov-additive jump process (N, J) is uniquely determined by the probabilities $P(s, t; y, A \times B)$. Since (N, J) is a Markov jump process, the infinitesimal transition rates

$$q(t; (x, y), A \times B) = \lim_{h \rightarrow 0} \frac{P(t, t+h; (x, y), A \times B) - 1_{A \times B}(x, y)}{h}$$

exist uniformly with respect to $(t, (x, y), A \times B)$ and the equality

$$q(t; (x, y), A \times B) = q(t; (0, y), (A - x) \times B) \quad (2.2)$$

follows from equation 2.1. Hence, we can define

$$q(t; y, A \times B) := q(t; (0, y), A \times B)$$

for all $t \in \mathbb{R}_0^+$, $y \in \Phi$ and $A \times B \in \mathcal{S}$, and as in definition A.2

$$\gamma(t, y) := -q(t; y, \{(0, y)\})$$

and

$$\gamma(t, y, A \times B) := q(t; y, (A \times B) \setminus \{(0, y)\})$$

as well as

$$p(t, y, A \times B) := \begin{cases} \frac{\gamma(t, y, A \times B)}{\gamma(t, y)} & \text{for } \gamma(t, y) > 0 \\ 1_{A \times B}(0, y) & \text{for } \gamma(t, y) = 0 \end{cases}$$

Finally, define the kernel $Q(t)$ on $\Sigma \times \Phi$ by

$$Q(t; (x, y), A \times B) := q(t; y, (A - x) \times B)$$

for all $t \in \mathbb{R}_0^+$, $(x, y) \in S$ and $A \times B \in \mathcal{S}$. The kernel $Q(t)$ is called the **generator** of (N, J) at time t .

Remark 2.3 By definition, a Markov-additive jump process is translation invariant or homogeneous in the first component. This leads to a self-similar structure of the generators $(Q(t) : t \in \mathbb{R}_0^+)$, which can be illustrated in the case of an homogeneous Markov-additive jump process with Φ being finite and $\Sigma = \mathbb{N}_0$ (the so-called Batch Markovian Arrival Process or BMAP, see Lucantoni [60]). Here, the generator Q , which is constant in time, takes the form

$$Q = \begin{pmatrix} D_0 & D_1 & D_2 & D_3 & \dots \\ & D_0 & D_1 & D_2 & \dots \\ & & D_0 & D_1 & \ddots \\ & & & D_0 & \ddots \\ & & & & \ddots \end{pmatrix}$$

with $m \times m$ -matrices $(D_n : n \in \mathbb{N}_0)$, m being the size of Φ . Omitting the first row and first column, one obtains the matrix Q again. This self-similarity yields many simplifications for the analysis of queues with BMAP arrivals. For a so-called level-dependent BMAP, which does not possess this self-similarity, and for the analysis of queues with level-dependent BMAP arrivals see Hofmann [50].

As in appendix A, the transition probabilities of the process will be derived as the solutions of the Kolmogorov differential equations via the method of successive approximations by Picard and Lindelöf.

Theorem 2.1 *Let (N, J) be a Markov-additive jump process. Then the Kolmogorov differential equations for (N, J) take the following form:*

$$\frac{\partial P(s, t; y, A \times B)}{\partial t} = \int_{\Sigma \times \Phi} q(t; w, (A - v) \times B) P(s, t; y, d(v, w))$$

2. Markov-Additive Jump Processes

for all $t > s$ (**Kolmogorov's forward equation**) and

$$\frac{\partial P(s, t; y, A \times B)}{\partial s} = - \int_{\Sigma \times \Phi} P(s, t; w, (A - v) \times B) q(s; y, d(v, w))$$

for all $s < t$ (**Kolmogorov's backward equation**).

Proof: This is immediate from theorem A.4 and equations 2.1 and 2.2.

Both differential equations contain a convolution in the first dimension. This is a consequence of the additivity which is defined on the marginal state space Σ . The convolution form will be preserved in the transition probabilities, as the next theorems show.

Theorem 2.2 *The transition probabilities of a Markov-additive jump process are uniquely determined by the infinitesimal rates $q(t; y, A \times B)$. They assume the form*

$$P(s, t; y, A \times B) = \sum_{n=0}^{\infty} P_b^{(n)}(s, t; y, A \times B)$$

with

$$P_b^{(0)}(s, t; y, A \times B) := e^{-\int_s^t \gamma(\tau, y) d\tau} 1_{A \times B}(0, y)$$

and recursively

$$P_b^{(n+1)}(s, t; y, A \times B) := \int_s^t \int_{\Sigma \times \Phi} P_b^{(n)}(\theta, t; w, (A - v) \times B) e^{-\int_s^\theta \gamma(\tau, y) d\tau} \gamma(\theta, y, d(v, w)) d\theta$$

for all $n \in \mathbb{N}_0$, using the notations from remark 2.2.

Proof: See theorem A.5 and equations 2.1 and 2.2.

If one develops this iteration formula, one obtains

$$\begin{aligned} P_b^{(n)}(s, t; y, A \times B) &= \\ &= \underbrace{\int_s^t \int_{u_1}^t \dots \int_{u_{n-1}}^t}_{n \text{ integrals}} \underbrace{\int_S \dots \int_S \int_{(A - \sum_{i=1}^{n-1} x_i) \times B}}_{n \text{ integrals}} e^{-\int_s^{u_1} \gamma(u; y) du} \gamma(u_1; y, d(x_1, y_1)) \dots \\ &\quad \dots e^{-\int_{u_{n-1}}^{u_n} \gamma(u; y_{n-1}) du} \gamma(u_n; y_{n-1}, d(x_n, y_n)) e^{-\int_{u_n}^t \gamma(u; y_n) du} du_1 \dots du_n \end{aligned}$$

as can be shown by induction. The transition probabilities can be computed iteratively by starting with

$$P_0(s, t; y, A \times B) := e^{-\int_s^t \gamma(\tau, y) d\tau} 1_{A \times B}(0, y)$$

and iterating

$$\begin{aligned} P_{n+1}(s, t; y, A \times B) &:= \\ &= \int_s^t \int_S P_n(\theta, t; w, (A - v) \times B) e^{-\int_s^\theta \gamma(\tau, y) d\tau} \gamma(\theta, y, d(v, w)) d\theta \\ &\quad + e^{-\int_s^t \gamma(\tau, y) d\tau} 1_{A \times B}(0, y) \end{aligned}$$

for all $n \in \mathbb{N}_0$ and $s < t \in \mathbb{R}_0^+$. Then we have $P(s, t) = \lim_{n \rightarrow \infty} P_n(s, t)$ for all $s < t \in \mathbb{R}_0^+$.

Theorem 2.3 *The transition probabilities of a Markov-additive jump process can also be expressed in the form*

$$P(s, t; y, A \times B) = \sum_{n=0}^{\infty} P^{(n)}(s, t; y, A \times B)$$

with

$$P^{(0)}(s, t; y, A \times B) := 1_{A \times B}(0, y)$$

and recursively

$$P^{(n+1)}(s, t; y, A \times B) := \int_s^t \int_{\Sigma \times \Phi} P^{(n)}(s, u; y, d(v, w)) q(u; w, (A - v) \times B) du$$

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

Proof: See theorem A.6 and equations 2.1 and 2.2.

Again by induction, one can prove that

$$P^{(n)}(s, t; y, A \times B) = \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} (Q(u_1) \dots Q(u_n)) ((0, y), A \times B) du_1 \dots du_n \quad (2.3)$$

with $Q(u)$ denoting the generator at time $u \in [s, t]$. An iteration for computing the transition probabilities is given by starting with

$$P_0(s, t; y, A \times B) := 1_{A \times B}(0, y)$$

and iterating by

$$P_{n+1}(s, t; y, A \times B) := \int_s^t \int_S P_n(s, u; y, d(v, w)) q(u; w, (A - v) \times B) du + 1_{A \times B}(0, y) \quad (2.4)$$

for all $n \in \mathbb{N}_0$. Then we have $P(s, t) = \lim_{n \rightarrow \infty} P_n(s, t)$ for all $s < t \in \mathbb{R}_0^+$. For the special case of a finite state space, this formula reduces to the iteration given in Bellman [19], p.168, or Kamke [53], p.52. In queueing theory, models with finite state spaces play an important role, e.g. for finite capacity systems (see section 4.1).

2. Markov-Additive Jump Processes

Remark 2.4 In order to illustrate the analogy to existing BMAP theory, define the following. For every $A \in \sigma(\Sigma)$ and $t \in \mathbb{R}_0^+$, define $D(t)_A(y, B) := Q(t; y, A \times B)$ for all $y \in \Phi$ and $B \in \sigma(\Phi)$. Then $D(t)_A$ is a kernel on Φ for every $A \in \sigma(\Sigma)$. Define the function $\Delta(t) : A \rightarrow D(t)_A$ on $\sigma(\Sigma)$ and write $\Delta(t)_A := \Delta(t)(A) = D(t)_A$. Further define a convolution of such functions by

$$(\Delta(u) * \Delta(v))_A(y, B) := \int_{s \in \Sigma} \int_{z \in \Phi} Q(u; y, ds \times dz) Q(v; z, (A - s) \times B)$$

for all $u, v \in \mathbb{R}_0^+$, $A \in \sigma(\Sigma)$, $y \in \Phi$ and $B \in \sigma(\Phi)$. Then also $(\Delta(u) * \Delta(v))_A$ is a kernel on Φ for every $A \in \sigma(\Sigma)$. By iteration, it is clear that

$$(\Delta(u_1) * \dots * \Delta(u_{n+1}))_A := ((\Delta(u_1) * \dots * \Delta(u_n)) * \Delta(u_{n+1}))_A$$

is again a kernel on Φ for every $A \in \sigma(\Sigma)$ and $n \in \mathbb{N}_0$. The identity kernel with respect to this convolution is given by $Id_A := 1_A(0) \cdot I$, with I denoting the identity kernel on Φ .

Now the transition probabilities from time s to time t can be expressed by

$$P(s, t) = \sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} (\Delta(u_1) * \dots * \Delta(u_n)) du_1 \dots du_n$$

meaning that for every $y \in \Phi$, $A \in \sigma(\Sigma)$ and $B \in \sigma(\Phi)$, we have

$$P(s, t; y, A \times B) = \left(\sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} (\Delta(u_1) * \dots * \Delta(u_n))_A du_1 \dots du_n \right) (y, B)$$

In the case of an homogeneous Markov-additive jump process, we can define $\Delta^0 := Id$ and iteratively $\Delta^{*(n+1)} := \Delta^{*n} * \Delta$. Then the above observation and theorem A.8 yield

$$P(s, t) = e^{*\Delta \cdot (t-s)} = \sum_{n=0}^{\infty} \frac{(t-s)^n}{n!} \Delta_A^{*n}$$

which for $\Sigma = \mathbb{N}_0$ and $\Phi = \{1, \dots, m\}$ reduces to the BMAP theory exposed in Baum [14].

2.2. Elementary Properties

Some elementary properties for Markov-additive jump processes can be taken from Çinlar [25] and Ezhov, Skorokhod [37], who examined the more general class of Markov-additive processes. This section contains some properties which mostly are immediate consequences of the definition 2.1. Obviously, the basic sample path properties of Markov jump processes hold for the subclass of Markov-additive jump processes, too. Furthermore, the distribution of the holding time in a state of the additive space Σ can be given.

Theorem 2.4 Let (N, J) be a Markov-additive jump process with state space $\Sigma \times \Phi$. Then the marginal process J is a Markov jump process with state space Φ and transition probabilities

$$P^\Phi(s, t; y, B) = P(s, t; y, \Sigma \times B)$$

for all $s < t \in \mathbb{R}_0^+$, $y \in \Phi$ and $B \in \sigma(\Phi)$.

Proof: This follows immediately from definitions 2.1 and A.1.

Remark 2.5 The above observation gives rise to a method of analysis which regards the marginal process J as the independent underlying Markov jump process that conditions the additive process N . The marginal state space Φ shall be called **phase space** of (N, J) , a single element of Φ is called a **phase**. The marginal process J shall be called **phase process** of (N, J) , while N shall be called **additive process** of (N, J) . The next theorem shows that N has conditionally independent increments given J . Furthermore, theorem 2.6 yields a representation of the generator as the product of two kernels.

Theorem 2.5 Let (N, J) be a Markov-additive jump process with state space $\Sigma \times \Phi$. Define $T := \inf\{t \in \mathbb{R}^+ : N_t \neq N_0\}$ as the holding time of the marginal process N in a state of Σ . Further, define the sub-Markov kernel $P(T > t)$ on Φ by

$$P(T > t)(y, B) := P(T > t, J_t \in B | J_0 = y)$$

for all $t \in \mathbb{R}^+$, $y \in \Phi$ and $B \in \sigma(\Phi)$. Then for all $s, t \in \mathbb{R}^+$,

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

and furthermore,

$$P(T > t)(y, B) = \sum_{n=0}^{\infty} \underbrace{\int_0^t \int_0^{u_n} \dots \int_0^{u_2}}_{n \text{ integrals}} (Q_0(u_1) \dots Q_0(u_n))(y, B) du_1 \dots du_n \quad (2.5)$$

with

$$Q_0(t)(y, B) := Q(t; y, \{0\} \times B)$$

for all $t \in \mathbb{R}_0^+$, $y \in \Phi$ and $B \in \sigma(\Phi)$.

Proof: The first equation follows from the definition of $P(T > t)$ and the fact that (N, J) is a Markov-additive jump process. For the second equation, note that the holding time T is identical to the first hitting time T_A of the set $A := \Sigma \setminus \{0\} \times \Phi$. Writing $A^c = \{0\} \times \Phi$, theorem A.19 yields that

$$\begin{aligned} P_y(T_A > t, J_t \in B) &= \\ &= \sum_{n=0}^{\infty} \underbrace{\int_0^t \int_{u_1}^t \dots \int_{u_{n-1}}^t}_{n \text{ integrals}} \underbrace{\int_{A^c} \dots \int_{A^c} \int_{\{0\} \times B}}_{n \text{ integrals}} e^{-\int_0^{u_1} \gamma(u; y) du} \gamma(u_1; y, d(0, y_1)) \dots \\ &\quad \dots e^{-\int_{u_{n-1}}^{u_n} \gamma(u; y_{n-1}) du} \gamma(u_n; y_{n-1}, d(0, y_n)) e^{-\int_{u_n}^t \gamma(u; y_n) du} du_1 \dots du_n \end{aligned}$$

2. Markov-Additive Jump Processes

for every $t \in \mathbb{R}^+$, initial state $(0, y) \in \Sigma \times \Phi$ and measurable set $B \in \sigma(\Phi)$. This equals the transition probability of a (sub-)Markov jump process with state space Φ and generator $(Q_0(t) : t \in \mathbb{R}_0^+)$. According to equation A.2, statement 2.5 is merely an alternative form for this transition probability.

☺

Remark 2.6 This result contains the statements of theorems 6.9 and 6.11 in Pacheco, Prabhu [77] as the special case of (N, J) being homogeneous, $\Sigma = \mathbb{R}_0^r$ and Φ being countable. This follows from theorem A.8.

Remark 2.7 The above theorem and formula 2.5 could serve to define a generalization of the phase-type or PH distribution (see Neuts [72], pp.231–248). The PH distribution is defined as the probability measure of the time until an homogeneous continuous-time Markov chain leaves a finite set of transient states. The generalization implied by formula 2.5 would yield the probability measure of the time until a general continuous-time Markov jump process leaves a general set of transient states. The exit rate vector in this generalization would be defined as $\eta(t, y) := -Q_0(t; y, S)$ and hence be dependent on time.

For later use, a fundamental theorem will be stated which asserts the existence of a representation of the infinitesimal transition rates (resp. the transition probabilities) as the product of a kernel on the phase space Φ (which is a function $\Phi \times \sigma(\Phi) \rightarrow [0, 1]$) and a kernel from Φ into the additive space Σ (which is a function $\Phi \times \sigma(\Sigma) \rightarrow [0, 1]$).

Theorem 2.6 For every $t \in \mathbb{R}_0^+$ and $y \in \Phi$, there is a kernel $K_{t,y} : \Phi \times \sigma(\Sigma) \rightarrow \mathbb{R}$ such that

$$Q(t; y, A \times B) = \int_B Q(t; y, \Sigma \times dz) K_{t,y}(z, A)$$

for all $A \in \sigma(\Sigma)$ and $B \in \sigma(\Phi)$. Furthermore, for every $s < t \in \mathbb{R}_0^+$ and $y \in \Phi$, there is a kernel $L_{s,t,y} : \Phi \times \sigma(\Sigma) \rightarrow \mathbb{R}$ such that

$$P(s, t; y, A \times B) = \int_B P(s, t; y, \Sigma \times dz) L_{s,t,y}(z, A)$$

for all $A \in \sigma(\Sigma)$ and $B \in \sigma(\Phi)$. The kernels $K_{t,y}$ and $L_{s,t,y}$ are almost surely uniquely determined.

Proof: See Bauer [13], p.397 (with $\mathcal{C} = \Sigma \times \sigma(\Phi)$), or Bourbaki [23], p.39 (with $p = pr_2$), since the charge $Q(t; y, \cdot)$ can be represented as the difference of two finite measures.

2.3. Markov-Additive Jump Processes on a Real Vector Space

Now assume that $(\Sigma, +)$ is a real vector space. Then $(\Sigma, +)$ has a base $(b_i : i \in I)$ with some index set I , and every element $x \in \Sigma$ has a unique representation $(c_i : i \in I)$ in \mathbb{R}^I with

respect to this base (cf. Nef [70], p.46). Denote the bijective mapping between an element $x \in \Sigma$ and its representation $(c_i : i \in I)$ by $f : \Sigma \rightarrow \mathbb{R}^I$. In order to simplify the notation in the following, an element $x \in \Sigma$ and its representation $f(x) \in \mathbb{R}^I$ shall be identified. Further, denote the projection from \mathbb{R}^I to the i -th component by $pr_i : \mathbb{R}^I \rightarrow \mathbb{R}$.

Under this assumption, transforms as well as expectations can be derived for the marginal process N of a Markov-additive jump processes (N, J) . Using the structure of the vector space, we can define the expectation of a Markov-additive jump process with state space $\Sigma \times \Phi$ as follows:

Definition 2.2 Let (N, J) be a Markov-additive jump process with state space $\Sigma \times \Phi$ and assume that $(\Sigma, +)$ is a real vector space. The **expectation** $E(N_t - N_s)$ of the marginal process N during the time interval $]s, t]$ is defined as the kernel on Φ with entries

$$E(N_t - N_s)(y, B) := (E(pr_i(N_t - N_s) \cdot 1_B(J_t) | J_s = y) : i \in I)$$

For real valued random variables, the Fourier transform has proved very useful, especially for determining moments (cf. Bauer [13], pp.183-223). The same analytical method shall be applied here for every dimension of the vector space Σ .

Definition 2.3 Let (N, J) be a Markov-additive jump process with state space $\Sigma \times \Phi$ and let $(\Sigma, +)$ be a real vector space. Then the **Fourier transform** $N_{s,t}^*$ of (N, J) over the time interval $]s, t]$ shall be defined as the function $r \rightarrow N_{s,t}^*(r)$ with

$$N_{s,t}^*(r)(y, B) := \left(\int_{-\infty}^{\infty} e^{i \cdot r c} P(s, t; y, pr_j^{-1}(dc) \times B) : j \in I \right)$$

for all $r \in \mathbb{R}$, $y \in \Phi$ and $B \in \sigma(\Phi)$, with i denoting the imaginary unit in the space \mathbb{C} of complex numbers.

Remark 2.8 Note that by theorem 2.6, the distribution $P(s, t; y, pr_j^{-1}(dc) \times B)$ in $c \in \mathbb{R}$ can be written as

$$P(s, t; y, pr_j^{-1}(dc) \times B) = \int_B P(s, t; y, \Sigma \times dz) L_{s,t,y}(z, dc)$$

with some kernel $L_{s,t,y}$ from Φ to \mathbb{R} . Hence the representation

$$pr_j(N_{s,t}^*(r)(y, B)) = \int_B \int_{-\infty}^{\infty} e^{i \cdot r c} L_{s,t,y}(z, dc) P(s, t; y, \Sigma \times dz)$$

holds and thus every component $pr_j(N_{s,t}^*(r))$ is a kernel on Φ .

Theorem 2.7 For every $j \in I$, $r \in \mathbb{R}$ and $t \in \mathbb{R}_0^+$, define the kernel $\varphi_{Q^{(j)}(t)}(r)$ on Φ by its entries

$$\begin{aligned} \varphi_{Q^{(j)}(t)}(r)(y, B) &:= \int_{-\infty}^{\infty} e^{i \cdot r c} Q(t; y, pr_j^{-1}(dc) \times B) \\ &= \int_B \int_{-\infty}^{\infty} e^{i \cdot r c} K_{t,y}(z, pr_j^{-1}(dc)) Q(t; y, \Sigma \times dz) \end{aligned}$$

2. Markov-Additive Jump Processes

for all $y \in \Phi$ and $B \in \sigma(\Phi)$, with some kernel $K_{t,y}$ from Φ to Σ (see theorem 2.6 for the second equality). Then the Fourier transform of (N, J) over the time interval $]s, t]$ can be written as

$$N_{s,t}^*(r) = \left(\sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} \varphi_{Q^{(j)}(u_1)}(r) \dots \varphi_{Q^{(j)}(u_n)}(r) du_1 \dots du_n : j \in I \right)$$

for all $r \in \mathbb{R}$, with the summand for $n = 0$ being the identity kernel I on Φ .

Proof: Fix $j \in I$ as well as $y \in \Phi$ and $B \in \sigma(\Phi)$. According to formula 2.3, the j th component of $N_{s,t}^*(r)(y, B)$ equals

$$\begin{aligned} pr_j(N_{s,t}^*(r)(y, B)) &= \sum_{n=0}^{\infty} \int_{-\infty}^{\infty} e^{i \cdot r c} P^{(n)}(s, t; y, pr_j^{-1}(dc) \times B) \\ &= \sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} \int_{-\infty}^{\infty} e^{i \cdot r c} (Q(u_1) \dots Q(u_n)) ((0, y), pr_j^{-1}(dc) \times B) du_1 \dots du_n \\ &= \sum_{n=0}^{\infty} \int_s^t \int_s^{u_n} \dots \int_s^{u_2} \int_{\mathbb{R} \times \Phi} Q(u_1; y, pr_j^{-1}(dc_1) \times dy_1) e^{i \cdot r c_1} \dots \\ &\quad \dots \int_{\mathbb{R} \times \Phi} Q(u_{n-1}; y_{n-2}, pr_j^{-1}(dc_{n-1}) \times dy_{n-1}) e^{i \cdot r c_{n-1}} \\ &\quad \int_{\mathbb{R} \times B} Q(u_n; y_{n-1}, pr_j^{-1}(dc_n) \times dy_n) e^{i \cdot r c_n} du_1 \dots du_n \end{aligned}$$

by writing out the convolution $c = c_1 + \dots + c_n$ explicitly. Further, application of theorem 2.6 leads to

$$\begin{aligned} pr_j(N_{s,t}^*(r)(y, B)) &= \\ &= \sum_{n=0}^{\infty} \int_s^t \int_s^{u_n} \dots \int_s^{u_2} \int_{\Phi} \int_{-\infty}^{\infty} e^{i \cdot r c_1} K_{u_1, y}(y_1, pr_j^{-1}(dc_1)) Q(u_1; y, \Sigma \times dy_1) \dots \\ &\quad \dots \int_{\Phi} \int_{-\infty}^{\infty} e^{i \cdot r c_{n-1}} K_{u_{n-1}, y_{n-2}}(y_{n-1}, pr_j^{-1}(dc_{n-1})) Q(u_{n-1}; y_{n-2}, \Sigma \times dy_{n-1}) \\ &\quad \int_B \int_{-\infty}^{\infty} e^{i \cdot r c_n} K_{u_n, y_{n-1}}(y_n, pr_j^{-1}(dc_n)) Q(u_n; y_{n-1}, \Sigma \times dy_n) du_1 \dots du_n \\ &= \sum_{n=0}^{\infty} \int_s^t \int_s^{u_n} \dots \int_s^{u_2} \int_{\Phi} \varphi_{Q^{(j)}(u_1)}(r)(y, dy_1) \dots \int_B \varphi_{Q^{(j)}(u_n)}(r)(y_{n-1}, dy_n) du_1 \dots du_n \\ &= \sum_{n=0}^{\infty} \int_s^t \int_s^{u_n} \dots \int_s^{u_2} (\varphi_{Q^{(j)}(u_1)}(r) \dots \varphi_{Q^{(j)}(u_n)}(r))(y, B) du_1 \dots du_n \end{aligned}$$

☺

Theorem 2.8 Define the kernel $M^{(j)}(t)$ on Φ by

$$M^{(j)}(t)(y, B) := \int_{-\infty}^{\infty} c Q(t; y, pr_j^{-1}(dc) \times B)$$

for all $t \in \mathbb{R}$, $j \in I$, $y \in \Phi$ and $B \in \sigma(\Phi)$. Then the expectation of the marginal process N over the time interval $]s, t]$ is given by

$$E(N_t - N_s) = \left(\int_s^t P^\Phi(s, u) M^{(j)}(u) P^\Phi(u, t) du : j \in I \right)$$

Proof: The expectation $E(N_t - N_s)$ can be computed as

$$E(N_t - N_s) = -i \cdot \frac{d}{dr} N_{s,t}^*(r) \Big|_{r=0}$$

recognizing that the differentiation can be performed component-wise. Then the j th component of $E(N_t - N_s)$ is

$$\begin{aligned} pr_j(E(N_t - N_s)) &= \\ &= -i \cdot \sum_{n=1}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \cdots \int_s^{u_2}}_{n \text{ integrals}} \frac{d}{dr} (\varphi_{Q^{(j)}(u_1)}(r) \cdots \varphi_{Q^{(j)}(u_n)}(r)) \Big|_{r=0} du_1 \cdots du_n \end{aligned}$$

according to the preceding theorem 2.7. Applying the product rule of differentiation, this equals

$$\begin{aligned} pr_j(E(N_t - N_s)) &= -i \cdot \sum_{n=1}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \cdots \int_s^{u_2}}_{n \text{ integrals}} \sum_{l=1}^n \varphi_{Q^{(j)}(u_1)}(0) \cdots \varphi_{Q^{(j)}(u_{l-1})}(0) \\ &\quad \left(\frac{d}{dr} \varphi_{Q^{(j)}(u_l)}(r) \Big|_{r=0} \right) \varphi_{Q^{(j)}(u_{l+1})}(0) \cdots \varphi_{Q^{(j)}(u_n)}(0) du_1 \cdots du_n \\ &= \sum_{l=1}^{\infty} \sum_{n=l}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \cdots \int_s^{u_{l+2}} \int_s^{u_{l+1}}}_{n-l \text{ integrals}} \underbrace{\int_s^{u_l} \cdots \int_s^{u_2}}_{l-1 \text{ integrals}} \varphi_{Q^{(j)}(u_1)}(0) \cdots \varphi_{Q^{(j)}(u_{l-1})}(0) du_1 \cdots du_{l-1} \\ &\quad \left(-i \cdot \frac{d}{dr} \varphi_{Q^{(j)}(u_l)}(r) \Big|_{r=0} \right) du_l \varphi_{Q^{(j)}(u_{l+1})}(0) \cdots \varphi_{Q^{(j)}(u_n)}(0) du_{l+1} \cdots du_n \end{aligned}$$

Acknowledging that

$$-i \cdot \frac{d}{dr} \varphi_{Q^{(j)}(u_l)}(r) \Big|_{r=0} (y, B) = \int_{-\infty}^{\infty} c Q(u_l; y, pr_j^{-1}(dc) \times B) = M^{(j)}(u_l)(y, B)$$

2. Markov-Additive Jump Processes

is the expectation kernel of the j th component and

$$\varphi_{Q^{(j)}(u)}(0)(y, B) = Q(u; y, \Sigma \times B)$$

is the infinitesimal transition rate of the phase process, leads to

$$\begin{aligned} pr_j(E(N_t - N_s)) &= \\ &= \int_s^t \left(\sum_{l=1}^{\infty} \underbrace{\int_s^{u_1} \cdots \int_s^{u_{l-1}}}_{l-1 \text{ integrals}} \varphi_{Q^{(j)}(u_1)}(0) \cdots \varphi_{Q^{(j)}(u_{l-1})}(0) du_1 \cdots du_{l-1} \right) M^{(j)}(u_l) \\ &\quad \left(\sum_{n=l}^{\infty} \underbrace{\int_{u_l}^t \int_{u_l}^{u_n} \cdots \int_{u_l}^{u_{l+2}}}_{n-l \text{ integrals}} \varphi_{Q^{(j)}(u_{l+1})}(0) \cdots \varphi_{Q^{(j)}(u_n)}(0) du_{l+1} \cdots du_n \right) du_l \\ &= \int_s^t P^\Phi(s, u_l) M^{(j)}(u_l) P^\Phi(u_l, t) du_l \end{aligned}$$

☺

Example 2.1 The previous results can be exemplified by the following process. Let Q^Φ be the generator of an homogeneous Markov jump process J with measurable state space $(\Phi, \sigma(\Phi))$. This will be the phase process. Assume that J has a stationary distribution π , i.e. $\pi Q^\Phi = 0$. Further, let σ be some probability distribution on $(\mathbb{R}, \mathcal{B})$ with \mathcal{B} denoting the Borel σ -algebra on \mathbb{R} . Specify $(\Sigma, +)$ to be the real line $(\mathbb{R}, +)$, i.e. Σ has dimension 1. Choosing a function $\lambda : \mathbb{R}_0^+ \times \Phi \rightarrow \mathbb{R}_0^+$ of time-dependent arrival rates with $\sup_{t \in \mathbb{R}_0^+, y \in \Phi} \lambda(t, y) < \infty$, we can define a Markov-additive jump process by

$$q(t; y, A \times B) := 1_A(0) \cdot Q^\Phi(y, B) + 1_B(y) \cdot \lambda(t, y)(\sigma(A) - 1_A(0))$$

for all $t \in \mathbb{R}_0^+$, $y \in \Phi$, $A \in \mathcal{B}$ and $B \in \sigma(\Phi)$. This process may be called a fluid compound MMPP with homogeneous phase process. The mean marginal rate of the additive process at time t and in phase y equals

$$M(t)(y, \Phi) = \int_{-\infty}^{\infty} c \cdot \lambda(t, y) d\sigma(c) = \lambda(t, y) \cdot E(\sigma)$$

with $E(\sigma)$ denoting the first moment of the distribution σ . For this process, the expectation of the marginal process on $\Sigma = \mathbb{R}$ over the time interval $]s, t]$ starting in phase equilibrium at time s equals

$$\begin{aligned} \int_{\Phi} \pi(dy) E(N_t - N_s)(y, \Phi) &= \int_{\Phi} \pi(dy) \int_s^t M(t)(y, \Phi) du \\ &= \int_s^t \int_{\Phi} \lambda(t, y) \pi(dy) du \cdot E(\sigma) \end{aligned}$$

according to theorem 2.8.

2.4. Markovian Arrival Processes

In queueing theory, the most important variable is that of the number of users or customers in the system (which is the queue or the queueing network). The values of this are limited to the set \mathbb{N}_0 of non-negative integers. Thus for most applications in queueing theory, the use of Markov-additive processes (N, J) on real vector spaces (of dimension $|I|$) with \mathbb{N}_0^I -valued marginal process N suffices.

If the support of the marginal process N is on \mathbb{N}_0^I , i.e. if $P(\text{pr}_i(N_t) \in \mathbb{N}_0) = 1$ for all $i \in I$ and $t \in \mathbb{R}_0^+$, then the easier notion of a z-transform serves the same purposes as the Fourier transform. In this section, the same analysis as in the preceding section 2.3 shall be carried out using z-transforms.

Definition 2.4 Let (N, J) be a Markov-additive jump process with state space $\Sigma \times \Phi$ and let $(\Sigma, +)$ be a real vector space having a base $(b_i : i \in I)$. If the marginal process N has support \mathbb{N}_0^I , then (N, J) shall be called **Markovian arrival process**.

Remark 2.9 The homogeneity of the first component implies that Markovian arrival processes have only non-negative increments in the first dimension. This follows from property 2.1, since for every $n \in \Sigma$, $k \in \mathbb{N}$ and any dimension $i \in I$ of the vector space,

$$P(s, t; (n, y), \text{pr}_i^{-1}(\text{pr}_i(n) - k) \times B) = P(s, t; y, \text{pr}_i^{-1}(-k) \times B) = 0$$

because of $\text{pr}_i^{-1}(-k) \notin \mathbb{N}_0^I$.

Definition 2.5 Let (N, J) be a Markovian arrival process. Define the **z-transform** of (N, J) over the time interval $]s, t]$ as the function $z \rightarrow N(s, t; z)$ with values being the kernels on Φ which are determined by

$$N(s, t; z)(y, B) := \left(\sum_{n=0}^{\infty} P(\text{pr}_i(N_t - N_s) = n, J_t \in B | J_s = y) z^n : i \in I \right)$$

for all $z \in \mathbb{C}$ with $|z| \leq 1$, $y \in \Phi$ and $B \in \sigma(\Phi)$.

Theorem 2.9 For every $i \in I$, $k \in \mathbb{N}_0$ and $t \in \mathbb{R}_0^+$, define the kernels $Q_k^{(i)}(t)$ on Φ by

$$Q_k^{(i)}(t)(y, B) := Q(t; y, \text{pr}_i^{-1}(k) \times B)$$

for all $y \in \Phi$ and $B \in \sigma(\Phi)$. Then the z-transform of (N, J) over the time interval $]s, t]$ can be written as

$$N(s, t; z) = \left(\sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} \sum_{k=0}^{\infty} Q_k^{(i)}(u_1) z^k \dots \sum_{k=0}^{\infty} Q_k^{(i)}(u_n) z^k du_1 \dots du_n : i \in I \right)$$

for all $z \in \mathbb{C}$ with $|z| \leq 1$.

2. Markov-Additive Jump Processes

Remark 2.10 Note that $\sum_{k=0}^{\infty} Q_k^{(i)}(u_j)z^k$ is a kernel on Φ for every $j \in \mathbb{N}$ and thus every product $\sum_{k=0}^{\infty} Q_k^{(i)}(u_1)z^k \dots \sum_{k=0}^{\infty} Q_k^{(i)}(u_n)z^k$ is a kernel on Φ , too.

Proof: Fix $i \in I$ as well as $y \in \Phi$ and $B \in \sigma(\Phi)$. According to formula 2.3, the i th component of $N(s, t; z)(y, B)$ equals

$$\begin{aligned}
 pr_i(N(s, t; z)(y, B)) &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} P^{(n)}(s, t; y, pr_i^{-1}(k) \times B) \cdot z^k \\
 &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} (Q(u_1) \dots Q(u_n)) ((0, y), pr_i^{-1}(k) \times B) du_1 \dots du_n \cdot z^k \\
 &= \sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} \sum_{k=0}^{\infty} \sum_{k_1 + \dots + k_n = k} \left(Q_{k_1}^{(i)}(u_1) \dots Q_{k_n}^{(i)}(u_n) \right) (y, B) \cdot z^k du_1 \dots du_n \\
 &= \sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} \left(\sum_{k=0}^{\infty} Q_k^{(i)}(u_1)z^k \dots \sum_{k=0}^{\infty} Q_k^{(i)}(u_n)z^k \right) (y, B) du_1 \dots du_n
 \end{aligned}$$

since the z-transform of the convolution of the kernels $Q_{k_1}^{(i)}(u_1), \dots, Q_{k_n}^{(i)}(u_n)$ equals the product of the z-transforms of $Q_{k_1}^{(i)}(u_1), \dots, Q_{k_n}^{(i)}(u_n)$.

☺

Theorem 2.10 The expectation of the marginal process N over the time interval $]s, t]$ is given by

$$E(N_t - N_s) = \left(\int_s^t P^\Phi(s, u) \sum_{k=1}^{\infty} k \cdot Q_k^{(i)}(u) P^\Phi(u, t) du : i \in I \right)$$

Proof: The expectation $E(N_t - N_s)$ can be computed as

$$E(N_t - N_s) = \left. \frac{d}{dz} N(s, t; z) \right|_{z=1}$$

recognizing that the differentiation can be performed component-wise. Then the i th component of $E(N_t - N_s)$ is

$$\begin{aligned}
 pr_i(E(N_t - N_s)) &= \\
 &= \sum_{n=1}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} \left. \frac{d}{dz} \left(\sum_{k=0}^{\infty} Q_k^{(i)}(u_1)z^k \dots \sum_{k=0}^{\infty} Q_k^{(i)}(u_n)z^k \right) \right|_{z=1} du_1 \dots du_n
 \end{aligned}$$

Denote the generator of the phase process at time u by $Q^\Phi(u)$ and note that

$$Q^\Phi(u) = \sum_{k=0}^{\infty} Q_k^{(i)}(u)$$

is independent of $i \in I$. Applying the product rule of differentiation yields

$$\begin{aligned} pr_i(E(N_t - N_s)) &= \\ &= \sum_{n=1}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} \sum_{l=1}^n Q^\Phi(u_1) \dots Q^\Phi(u_{l-1}) \left(\sum_{k=1}^{\infty} k Q_k^{(i)}(u_l) \right) Q^\Phi(u_{l+1}) \dots Q^\Phi(u_n) \\ &\quad du_1 \dots du_n \\ &= \sum_{l=1}^{\infty} \sum_{n=l}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_{l+2}} \int_s^{u_{l+1}} \int_s^{u_l} \dots \int_s^{u_2}}_{n-l \text{ integrals}} \underbrace{Q^\Phi(u_1) \dots Q^\Phi(u_{l-1})}_{l-1 \text{ integrals}} du_1 \dots du_{l-1} \\ &\quad \left(\sum_{k=1}^{\infty} k Q_k^{(i)}(u_l) \right) du_l Q^\Phi(u_{l+1}) \dots Q^\Phi(u_n) du_{l+1} \dots du_n \\ &= \int_s^t \left(\sum_{l=1}^{\infty} \underbrace{\int_s^{u_l} \dots \int_s^{u_2}}_{l-1 \text{ integrals}} Q^\Phi(u_1) \dots Q^\Phi(u_{l-1}) du_1 \dots du_{l-1} \right) \sum_{k=1}^{\infty} k Q_k^{(i)}(u_l) \\ &\quad \left(\sum_{n=l}^{\infty} \underbrace{\int_{u_l}^t \int_{u_l}^{u_n} \dots \int_{u_l}^{u_{l+2}}}_{n-l \text{ integrals}} Q^\Phi(u_{l+1}) \dots Q^\Phi(u_n) du_{l+1} \dots du_n \right) du_l \\ &= \int_s^t P^\Phi(s, u_l) \left(\sum_{k=1}^{\infty} k Q_k^{(i)}(u_l) \right) P^\Phi(u_l, t) du_l \end{aligned}$$

remembering that the transition probabilities of the phase process are denoted by $P^\Phi(s, t)$.
 ☺

Remark 2.11 The first moment formulae given in Pacheco, Prabhu [77], theorem 6.15 and corollary 6.16(a), can be obtained from this theorem as the special case of (N, J) being homogeneous, $\Sigma = \mathbb{N}_0^r$ and Φ being countable.

2.5. Laws of Large Numbers

For some special cases of Markov-additive jump processes on real vector spaces, the asymptotic behaviour can be described in terms of strong laws of large numbers. First, this will be

2. Markov-Additive Jump Processes

done for processes with periodic generators (cf. appendix A.3). The case of homogeneous processes shall turn out to be a corollary of the periodic case. In this section, convergence on the vector space Σ shall be defined in a weak sense as convergence in every dimension of Σ .

Theorem 2.11 *Let (N, J) denote a periodic Markov-additive jump process with period T and be $(\Sigma, +)$ a real vector space with base $(b_i : i \in I)$. Let the phase process J have a periodic family of asymptotic distributions $(\pi_t : t \in [0, T])$ as defined in appendix A.3. Define the mean rate vector during a period length by*

$$\lambda := \left(\frac{1}{T} \int_0^T \int_{\Phi} d\pi_t(y) \int_{-\infty}^{\infty} c Q(t; y, pr_i^{-1}(dc) \times \Phi) dt : i \in I \right)$$

Assume that $\|pr_i(E(N_t))\| < M(t) < \infty$ for all $i \in I$ and $t \in [0, T]$. Then

$$\frac{N_t}{t} \rightarrow \lambda \quad \text{as } t \rightarrow \infty$$

P -almost surely for all initial distributions.

Proof: Because of property 2.1, the evolution of the process (N, J) depends only on the initial distribution μ of J_0 on the phase space Φ . Since further for every initial value $N_0 \neq 0$ the quotient N_0/t vanishes for $t \rightarrow \infty$, we can assume $N_0 = 0$ almost surely without loss of generality. Let N_t^μ denote the marginal process N at time t under initial phase distribution μ . Further, let $\varphi_t = \mu P^\Phi(0, t)$ denote the distribution of the phase process at time t .

Choose any $\varepsilon > 0$. Then there is a number $k \in \mathbb{N}$ such that $\|\mu P^\Phi(0, nT) - \pi_0\| < \varepsilon$ for all $n \geq k$ and the representation

$$\begin{aligned} \frac{N_t}{t} &= \frac{1}{t} (N_{kT} + (N_{\lfloor t/T \rfloor T} - N_{kT}) + (N_t - N_{\lfloor t/T \rfloor T})) \\ &= \frac{1}{t} N_{kT} + \frac{1}{t} N_{(\lfloor t/T \rfloor - k)T}^{\varphi_{kT}} + \frac{1}{t} (N_t - N_{\lfloor t/T \rfloor T}) \end{aligned}$$

holds because of the periodicity of the process. Since the expectation kernel $E(N_s)$ is bounded for all $s \in \mathbb{R}_0$, the first and the last term will vanish for $t \rightarrow \infty$ almost surely. The second term equals

$$\frac{1}{t} N_{(\lfloor t/T \rfloor - k)T}^{\varphi_{kT}} = \frac{1}{t} N_{(\lfloor t/T \rfloor - k)T}^{\pi_0} + \frac{1}{t} N_{(\lfloor t/T \rfloor - k)T}^{\varphi_{kT} - \pi_0} = \frac{1}{t} \sum_{m=k}^{\lfloor t/T \rfloor - 1} N_T^{\pi_0} + \frac{1}{t} N_{(\lfloor t/T \rfloor - k)T}^{\varphi_{kT} - \pi_0}$$

since $\pi_0 = \pi_0 P^\Phi(0, T)$ is invariant. The first term of this sum tends to

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{m=k}^{\lfloor t/T \rfloor - 1} N_T^{\pi_0} &= \lim_{t \rightarrow \infty} \frac{(\lfloor t/T \rfloor - k)T}{t} \frac{1}{(\lfloor t/T \rfloor - k)T} \sum_{m=k+1}^{\lfloor t/T \rfloor} N_T^{\pi_0} \\ &= \lim_{t \rightarrow \infty} \frac{(\lfloor t/T \rfloor - k)T}{t} \cdot \lim_{n \rightarrow \infty} \frac{1}{(n - k)T} \sum_{m=k+1}^n N_T^{\pi_0} \\ &= \frac{1}{T} E(N_T^{\pi_0}) = \lambda \end{aligned}$$

almost surely, according to Kolmogorov's law of large numbers (cf. Bauer [13], p.86) and theorem 2.8. The assumption $\|\mu P^\Phi(0, kT) - \pi_0\| < \varepsilon$ implies

$$P\left(N_{(\lfloor t/T \rfloor - k)T}^{\varphi_{kT}} \neq N_{(\lfloor t/T \rfloor - k)T}^{\pi_0}\right) < \varepsilon$$

which means that

$$N_{(\lfloor t/T \rfloor - k)T}^{\varphi_{kT} - \pi_0} = N_{(\lfloor t/T \rfloor - k)T}^{\varphi_{kT}} - N_{(\lfloor t/T \rfloor - k)T}^{\pi_0} = 0$$

with probability $1 - \varepsilon$. Hence with probability $1 - \varepsilon$, the convergence $\frac{N_t}{t} \rightarrow \lambda$ holds. Since ε can be chosen arbitrarily small, there can be no set N with $P(N) > 0$ such that $\frac{N_t(\omega)}{t} \not\rightarrow \lambda$ for all $\omega \in N$.

☺

Remark 2.12 If there is a norm on Σ (e.g. the supremum norm $\|x\| := \sup_{i \in I} |pr_i(x)|$ for all $x \in \Sigma$) and Σ is complete with respect to this norm, i.e. if Σ is a Banach space, then the above theorem is valid for convergence in terms of the norm on Σ , too. This can be proven in exactly the same way by using the strong law of large numbers on Banach spaces (cf. Mourier [68] or Beck [18]) instead of Kolmogorov's law of large numbers, which is a statement for real-valued random variables.

Theorem 2.12 Let (N, J) denote an homogeneous Markov-additive jump process and be $(\Sigma, +)$ a real vector space with base $(b_i : i \in I)$. Assume that the phase process J has an asymptotic distribution π . Define the mean rate vector by

$$\lambda := \left(\int_{\Phi} d\pi(y) \int_{-\infty}^{\infty} c Q(y, pr_i^{-1}(dc) \times \Phi) : i \in I \right)$$

If $\|pr_i(E(N_1))\| < M < \infty$ for all $i \in I$, then

$$\frac{N_t}{t} \rightarrow \lambda \quad \text{as } t \rightarrow \infty$$

P -almost surely for all initial distributions.

Proof: This follows immediately from theorem 2.11, since an homogeneous process with asymptotic distribution π is a periodic process with arbitrary period length $T > 0$ and periodic family $(\pi_t = \pi : t \in [0, T[)$ of asymptotic distributions. Furthermore,

$$\int_{\Phi} d\pi(y) \int_{-\infty}^{\infty} c Q(y, pr_i^{-1}(dc) \times \Phi) = \frac{1}{T} \int_0^T \int_{\Phi} d\pi_t(y) \int_{-\infty}^{\infty} c Q(t; y, pr_i^{-1}(dc) \times \Phi) dt$$

holds for all $i \in I$, since $Q(t; y, pr_i^{-1}(dc) \times \Phi) = Q(y, pr_i^{-1}(dc) \times \Phi)$ is constant in t .

☺

Remark 2.13 The statement of this theorem is the same as the result in Pacheco, Prabhu [77], theorem 6.17, if one specifies $\Sigma = IN_0^r$ and Φ being countable.

2. *Markov-Additive Jump Processes*

3. Spatial Markovian Arrival Processes

The recent rise of mobile communication systems gave reason to the development of queueing models which are capable of incorporating spatial features. Introducing spatial arrival processes is of great help for modelling mobile communication networks, since users can be characterized more adequately by properties depending on their position. To this end, the concept of a time-space process is needed, which is able to describe the time dynamics of arrivals as well as their distribution in space.

In this chapter, the concept of Batch Markovian Arrival Processes (or shortly BMAPs, see Lucantoni [60]) will be generalized towards a class of such time-space processes, which shall be called Spatial Markovian Arrival Processes (or shortly SMAPs). This generalization points in three directions. First, the phase space is allowed to be general. Second, the generator of an SMAP may depend on time, i.e. SMAPs provide for time-inhomogeneous behaviour. And finally, the attribute "spatial" appearing in the name stems from the feature that arrivals may assume a location in some space. Thus they are distributed not only with respect to time but also with respect to their location in a so-called arrival space. This is important for modelling mobile communication networks, for which the spatial distribution of users is of relevance and often the arrival intensities and service requirements may depend on the location a user calls from. Furthermore, spatial distributions of users typically will include spatial correlations among the different locations of arrivals in a batch. The spatial features of SMAPs are inspired by and based upon an earlier concept of Markovian spatial arrival processes introduced by Baum [16]. SMAPs keep the most important structural properties of BMAPs, such that the computations necessary for analyzing them and the respective queues are still tractable.

In section 3.1, the class of spatial Markovian arrival processes is defined and constructed from basic ingredients. The most important properties of SMAPs are derived in section 3.2. In particular, the transition probabilities of SMAPs are specified and a law of large numbers is given for periodic SMAPs. The last section contains some simple examples.

3.1. Definition and Construction

In this section, the so-called spatial Markovian arrival process (or shortly SMAP) will be introduced. This shall be done by using the concept of stochastic point fields developed in

3. Spatial Markovian Arrival Processes

appendix B. The space of counting functions on some arrival space will take the place of the additive space of a Markovian arrival process as defined in definition 2.4. Thus, point fields serve as a description for the spatial distribution of arrivals for SMAPs and henceforth, they shall be called **arrival fields**.

After defining spatial Markovian arrival processes as a special case of Markovian arrival processes, a way of constructing SMAPs from a given generator of the phase process and a set of finite-dimensional marginal distributions of the arrival fields is described.

Definition 3.1 Let R be a Polish space and $\sigma(R)$ the Borel σ -algebra on R . Further, let \mathcal{C} denote the space of counting functions on $(R, \sigma(R))$. Let $\sigma(\mathcal{C})$ be the σ -algebra on \mathcal{C} which is constructed in theorem B.3. Further, let Φ denote a separable metric space, which is locally compact, and let $\sigma(\Phi)$ denote a σ -algebra on Φ . A Markov-additive jump process (N, J) with state space $S := \mathcal{C} \times \Phi$ shall be called **spatial Markovian arrival process** (or shortly **SMAP**) with **arrival space** $(R, \sigma(R))$.

Remark 3.1 Since $\mathcal{C} \subset \mathbb{N}_0^{\sigma(R)}$, a spatial Markovian arrival process (N, J) is a special kind of Markovian arrival processes as defined in definition 2.4. Hence we can call Φ the phase space and J the phase process of (N, J) .

Remark 3.2 The batch Markovian arrival process (or shortly BMAP) developed by Lucantoni [60] is an homogeneous SMAP with a trivial arrival space $R = \{q\}$ consisting of only one element and a finite phase space $\Phi = \{1, \dots, m\}$. Hence, all results for SMAPs and queues with SMAPs are valid for BMAPs and queues with BMAPs, too. This observation will be illustrated in sections 4.2 and 6.4.

The remaining results show a way of constructing an SMAP. First, conversely to theorem 2.6, a generator on the state space $\mathcal{C} \times \Phi$ is composed of a kernel on Φ and a kernel from Φ to $\sigma(\mathcal{C})$. After that, it is shown that the second kernel (which includes a measure on the space \mathcal{C} of counting functions) can be determined uniquely by a sufficiently large set of finite-dimensional marginal distributions.

Theorem 3.1 Let $D^\Phi(t)$ be the (time-dependent) generator of a Markov jump process on Φ . Further, let $K_{t,y} : \Phi \times \sigma(\mathcal{C}) \rightarrow [0, 1]$ be a Markov kernel for every $y \in \Phi$ and $t \in \mathbb{R}_0^+$. Define

$$D(t; y, A \times B) := \int_B K_{t,y}(z, A) D^\Phi(t; y, dz)$$

for all $t \in \mathbb{R}_0^+$, $y \in \Phi$, $A \in \sigma(\mathcal{C})$ and $B \in \sigma(\Phi)$. Then $D(t)$ is the (time-dependent) generator of a spatial Markovian arrival process.

Proof: Define the same norm on the space of kernels on S as in remark A.1, namely

$$\|K\| := \sup_{x \in S, A \in \mathcal{S}} |K(x, A)|$$

for every kernel K on S . Since $D^\Phi(t)$ is a generator, the function $t \rightarrow D^\Phi(t)$ is bounded by theorem A.1. Because $K_{t,y}$ is a Markov kernel for every $y \in \Phi$ and $t \in \mathbb{R}_0^+$, this implies that the function $t \rightarrow D(t)$ is bounded, too.

Fix any $t \in \mathbb{R}_0^+$. Then

$$\begin{aligned} \|D(t+h) - D(t)\| &= \sup_{y \in S, A \times B \in \mathcal{S}} |D(t+h; y, A \times B) - D(t; y, A \times B)| \\ &\leq \sup_{y \in S, B \in \sigma(\Phi)} |D^\Phi(t+h; y, B) - D^\Phi(t; y, B)| \end{aligned}$$

since $K_{t,y}$ is a Markov kernel, i.e. $K_{t,y}(z, A) \in [0, 1]$ for all $z \in \Phi$ and $A \in \sigma(\mathcal{C})$. Since the function $t \rightarrow D^\Phi(t)$ is continuous, we have

$$\|D(t+h) - D(t)\| \leq \|D^\Phi(t+h) - D^\Phi(t)\| \rightarrow 0$$

as $h \rightarrow 0$. Hence, the function $t \rightarrow D(t)$ is continuous.

Now the statement follows by remark A.8.

☺

Remark 3.3 According to definition A.2 and section A.1, a generator D^Φ can be constructed by choosing values $\gamma(t, y) \in \mathbb{R}^+$ for all $y \in \Phi$ and $t \in \mathbb{R}_0^+$ and jump kernels $p(t; \cdot, \cdot) : \Phi \times \sigma(\Phi) \rightarrow [0, 1]$ for all $t \in \mathbb{R}_0^+$ such that the conditions in definition A.7 are fulfilled.

In order to uniquely determine the measure part $K_{t,y}(z, \cdot)$ of the kernel $K_{t,y}$, it suffices to specify the finite-dimensional marginal measures.

Definition 3.2 For $n \in \mathbb{N}$, $S_1, \dots, S_n \in \sigma(R)$ and $k_1, \dots, k_n \in \mathbb{N}_0$, define

$$pr_{(S_1, \dots, S_n)}^{-1}(k_1, \dots, k_n) := \{N \in \mathcal{C} : N(S_i) = k_i \forall i \in \{1, \dots, n\}\}$$

as the inverse image of the projection of functions in \mathcal{C} to values at (S_1, \dots, S_n) . Further, denote the stochastic point fields

$$\Pi_{t,y,z} := K_{t,y}(z, \cdot)$$

and their finite-dimensional marginal distributions at (S_1, \dots, S_n) by

$$\Pi_{t,y,z}(S_1, \dots, S_n)(k_1, \dots, k_n) := \Pi_{t,y,z}(pr_{(S_1, \dots, S_n)}^{-1}(k_1, \dots, k_n))$$

for all $t \in \mathbb{R}_0^+$ and $y, z \in \Phi$.

Theorem 3.2 An SMAP (N, J) on $(R, \sigma(R))$ is uniquely determined by the generator $D^\Phi(t)$ of its phase process and the marginal distributions $\Pi_{t,y,z}(S_1, \dots, S_n)$ of its arrival fields on finite families of disjoint subsets $S_1, \dots, S_n \in \sigma(R)$ from a semi-ring of bounded sets generating $\sigma(R)$.

Proof: According to theorem B.4, the measure $\Pi_{t,y,z} = K_{t,y}(z, \cdot)$ is uniquely determined by its marginal measures $\Pi_{t,y,z}(S_1, \dots, S_n)$ on finite families of disjoint subsets $S_1, \dots, S_n \in \sigma(R)$ from a semi-ring of bounded sets generating $\sigma(R)$. With this, the statement follows from theorem 3.1.

☺

3.2. Properties

In this section, the most important properties of SMAPs shall be derived. It turns out that it suffices to consider a sufficiently large set of finite-dimensional marginal processes of an SMAP, in order to describe the whole process. Formulae for the transition probabilities, which uniquely determine an SMAP, are taken from the more general concept of Markovian arrival processes and specified for the case of SMAPs. As for Markovian arrival processes, the concept of z -transforms leads to an expression for the expectation kernel of an SMAP. For periodic and homogeneous SMAPs, strong laws of large numbers describe their asymptotic behaviour.

3.2.1. Marginal Processes

The state space of the marginal process N is the space \mathcal{C} of counting functions. Since a probability measure on \mathcal{C} is uniquely determined by a sufficiently large set of finite-dimensional marginal distributions (see theorem B.4), it is possible to uniquely determine N by a set of finite-dimensional marginal processes of N .

Definition 3.3 Define the stochastic processes

$$N(S_1, \dots, S_n) := pr_{(S_1, \dots, S_n)}(N)$$

for all $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$. Then the process $(N(S_1, \dots, S_n), J)$ shall be called the **marginal process** of (N, J) on (S_1, \dots, S_n) .

Theorem 3.3 For $t \in \mathbb{R}_0^+$, $n \in \mathbb{N}$, $S_1, \dots, S_n \in \sigma(R)$ and $k_1, \dots, k_n \in \mathbb{N}_0$, define the kernels $D_{k_1, \dots, k_n}(t; S_1, \dots, S_n)$ on Φ by

$$D_{k_1, \dots, k_n}(t; S_1, \dots, S_n)(y, A) := D(t; y, pr_{(S_1, \dots, S_n)}^{-1}(k_1, \dots, k_n) \times A)$$

for all $y \in \Phi$ and $A \in \sigma(\Phi)$. Then the sequence

$$\Delta(t; S_1, \dots, S_n) := (D_{k_1, \dots, k_n}(t; S_1, \dots, S_n) : k_1, \dots, k_n \in \mathbb{N}_0)$$

uniquely determines the generator of the marginal process $(N(S_1, \dots, S_n), J)$ and thus the process $(N(S_1, \dots, S_n), J)$ itself.

Proof: This follows from the definitions of $D_{k_1, \dots, k_n}(t; S_1, \dots, S_n)$ and $N(S_1, \dots, S_n)$.

Definition 3.4 The time-dependent sequence $\Delta(t; S_1, \dots, S_n)$ shall be called **generating sequence** of the marginal process $(N(S_1, \dots, S_n), J)$.

Theorem 3.4 An SMAP (N, J) is uniquely determined by the time-dependent generating sequences $\Delta(t; S_1, \dots, S_n)$ of the marginal processes $(N(S_1, \dots, S_n), J)$ on all finite families of disjoint subsets $S_1, \dots, S_n \in \sigma(R)$ from a semi-ring of bounded sets generating $\sigma(R)$.

Proof: This follows from theorems 3.3 and B.4.

3.2.2. Transition Probabilities

In order to determine the transition probabilities of an SMAP (N, J) , a description of distributions on the space \mathcal{C} by means of finite-dimensional marginal distributions is helpful again. In this subsection, the transition probabilities of (N, J) are given in terms of the transition probabilities of the marginal processes $(N(S_1, \dots, S_n), J)$ for any $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$. This method further yields explicit formulae to compute those transition probabilities via convolutions of kernels on Φ which are indexed within \mathbb{N}_0^n for some $n \in \mathbb{N}$.

Definition 3.5 Define the kernel $N(s, t; C)$ on Φ by

$$N(s, t; C)(y, A) := P(s, t; y, C \times A)$$

for every $y \in \Phi$ and $A \in \sigma(\Phi)$. Further define for any $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$

$$N_{k_1, \dots, k_n}(s, t; S_1, \dots, S_n) := N(s, t; pr_{(S_1, \dots, S_n)}^{-1}(k_1, \dots, k_n))$$

as a kernel on Φ .

Remark 3.4 Note that by property 2.1, the kernels $(N(s, t; \cdot) : s < t \in \mathbb{R}_0^+)$ uniquely determine all transition probabilities of (N, J) . Furthermore, these are already determined by the transition probability kernels $N(s, t; S_1, \dots, S_n)$ on finite families of disjoint subsets $S_1, \dots, S_n \in \sigma(R)$ from a semi-ring of bounded sets generating $\sigma(R)$, as theorem B.4 implies.

In the case of finite-dimensional marginal processes of an SMAP, the convolution formulae for the transition probabilities, which are derived as the solutions of the Kolmogorov forward equations, take the following form:

Definition 3.6 For all $s < t \in \mathbb{R}_0^+$, $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$ set

$$N_{i_1, \dots, i_n}^{(0)}(s, t; S_1, \dots, S_n)(y, A) := \begin{cases} 1 & \text{for } \sum_{m=1}^n i_m = 0, y \in A \\ 0 & \text{otherwise} \end{cases}$$

for $i_1, \dots, i_n \in \mathbb{N}_0$, $y \in \Phi$ and $A \in \sigma(\Phi)$. Further, define the sequence

$$N^{(1)}(s, t; S_1, \dots, S_n) := \int_s^t \Delta(u; S_1, \dots, S_n) du$$

and recursively for $m \in \mathbb{N}$

$$N_{i_1, \dots, i_n}^{(m)}(s, t; S_1, \dots, S_n) :=$$

$$\sum_{j_1=0}^{i_1} \dots \sum_{j_n=0}^{i_n} \int_s^t N_{j_1, \dots, j_n}^{(m-1)}(s, u; S_1, \dots, S_n) D_{i_1-j_1, \dots, i_n-j_n}(u; S_1, \dots, S_n) du$$

for all $i_1, \dots, i_n \in \mathbb{N}_0$.

3. Spatial Markovian Arrival Processes

Theorem 3.5 For every $s < t \in \mathbb{R}_0^+$, the transition probability kernel $N(s, t) := N(s, t; \cdot)$ is given by

$$N(s, t) = \sum_{m=0}^{\infty} N^{(m)}(s, t),$$

meaning that for $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$

$$N_{k_1, \dots, k_n}(s, t; S_1, \dots, S_n) = \sum_{m=0}^{\infty} N_{k_1, \dots, k_n}^{(m)}(s, t; S_1, \dots, S_n)$$

for all $k_1, \dots, k_n \in \mathbb{N}_0$.

Proof: Since SMAPs are special Markov-additive jump processes, the transition probabilities are given in theorem 2.3. For the case of finite-dimensional marginal processes of an SMAP, they assume the above form.

Theorem 3.6 An iteration formula for the computation of the transition probabilities $N(s, t; S_1, \dots, S_n)(y, A)$ is given by starting with

$$\begin{aligned} N_{i_1, \dots, i_n}^{[0]}(s, t; S_1, \dots, S_n)(y, A) &:= 1_{\{0\} \times A} \left(\sum_{m=1}^n i_m, y \right) \\ &= \begin{cases} 1 & \text{for } \sum_{m=1}^n i_m = 0, y \in A \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

and iterating by

$$\begin{aligned} N_{i_1, \dots, i_n}^{[m]}(s, t; S_1, \dots, S_n) &:= \\ &= \sum_{j_1=0}^{i_1} \dots \sum_{j_n=0}^{i_n} \int_s^t N_{j_1, \dots, j_n}^{[m-1]}(s, u; S_1, \dots, S_n) D_{i_1-j_1, \dots, i_n-j_n}(u; S_1, \dots, S_n) du \\ &\quad + 1_{\{0\} \times A} \left(\sum_{m=1}^n i_m, y \right) \end{aligned}$$

for all $i_1, \dots, i_n \in \mathbb{N}_0$ and $m \in \mathbb{N}$.

Proof: This is merely the form of the iteration formula 2.4 for this special case.

Remark 3.5 For the special case of a finite state space, this formula reduces to the iteration given in Bellman [19], p.168, or Kamke [53], p.52.

In the case of homogeneous SMAPs, the above formulae assume a somewhat simpler form. For a special class of homogeneous SMAPs with finite phase space, this form has been derived in Baum [14, 16].

Let (N, J) denote an homogeneous SMAP with arrival space $(R, \sigma(R))$. Then for all $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$, the generating sequences

$$\Delta(S_1, \dots, S_n) := \Delta(t; S_1, \dots, S_n) = (D_{k_1, \dots, k_n}(S_1, \dots, S_n) : (k_1, \dots, k_n) \in (\mathbb{N}_0)^n)$$

are constant in time $t \in \mathbb{R}_0^+$.

Definition 3.7 For all $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$ set

$$\Delta(S_1, \dots, S_n)_{k_1, \dots, k_n}^{*0}(y, A) := \begin{cases} 1 & \text{for } \sum_{i=1}^n k_i = 0, y \in A \\ 0 & \text{otherwise} \end{cases}$$

for $k_1, \dots, k_n \in \mathbb{N}_0, y \in \Phi$ and $A \in \sigma(\Phi)$. Further, define the sequence

$$\Delta(S_1, \dots, S_n)^{*1} := \Delta(S_1, \dots, S_n)$$

and recursively for $m \in \mathbb{N}$

$$\Delta(S_1, \dots, S_n)_{k_1, \dots, k_n}^{*m} := \sum_{l_1=0}^{k_1} \dots \sum_{l_n=0}^{k_n} \Delta(S_1, \dots, S_n)_{l_1, \dots, l_n}^{*m-1} D_{k_1-l_1, \dots, k_n-l_n}(S_1, \dots, S_n)$$

Theorem 3.7 For an homogeneous SMAP, the finite-dimensional transition probability kernels $N(s, t; S_1, \dots, S_n)$ can be written as

$$N(s, t; S_1, \dots, S_n) = e^{*\Delta(S_1, \dots, S_n) \cdot (t-s)},$$

which means that the probability of k_i arrivals in S_i (for $i = 1, \dots, n$) during the time interval $]s, t]$ is

$$N_{k_1, \dots, k_n}(s, t; S_1, \dots, S_n) = \sum_{m=0}^{\infty} \frac{(t-s)^m}{m!} \Delta(S_1, \dots, S_n)_{k_1, \dots, k_n}^{*m}$$

with the above definition 3.7 of the convolution operator $*$.

Proof: The formulae in theorem 3.5 can be reduced to this exponential form for the homogeneous case, as is shown in theorem A.8.

Remark 3.6 In the homogeneous case, a computation of the transition probabilities can be pursued by solving the Kolmogorov forward differential equations

$$\frac{\partial}{\partial t} N(s, t; S_1, \dots, S_n) = N(s, t; S_1, \dots, S_n) * \Delta(S_1, \dots, S_n)$$

with initial value $N(s, s; S_1, \dots, S_n) = I$ via the Runge–Kutta method. This is possible in a straightforward way without encountering numerical problems (cf. Asmussen [6], Gilbert [43] or Moler, van Loan [67]).

3. Spatial Markovian Arrival Processes

Example 3.1 Assume that there is a function $\lambda : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ such that $\Delta(t) = \lambda(t) \cdot \Delta$ for all $t \in \mathbb{R}_0^+$. Then the transition probability kernel from time $s \in \mathbb{R}_0^+$ to time $t > s$ can be simplified to the form

$$N(s, t) = e^{*\Delta \cdot \int_s^t \lambda(u) du}$$

according to remark A.9 and theorem A.21. This means that the probability of k_i arrivals in $S_i \in \sigma(R)$ (for $i = 1, \dots, n$) during the time interval $]s, t]$ can be written as

$$N_{k_1, \dots, k_n}(s, t; S_1, \dots, S_n) = \sum_{m=0}^{\infty} \frac{\left(\int_s^t \lambda(u) du \right)^m}{m!} \Delta(S_1, \dots, S_n)_{k_1, \dots, k_n}^{*m}$$

using definition 3.7 for the convolution operator $*$.

As a useful corollary, which will be important for the estimation procedure in chapter 6, we get

Theorem 3.8 For an homogeneous SMAP with finite phase space, the distribution of the interarrival times on the finite family (S_1, \dots, S_n) of sets is phase-type and determined by

$$N_{0, \dots, 0}(s, t; S_1, \dots, S_n) = e^{D_{0, \dots, 0}(S_1, \dots, S_n)(t-s)}$$

Proof: The formula follows immediately from the above theorem 3.7 and the definition 3.7 of the convolution operator $*$. If the distribution of the phases at time s is given by the probability vector π , then it is easy to verify that the time duration from s until the first arrival in (S_1, \dots, S_n) after s has phase-type distribution with representation $(\pi, D_{0, \dots, 0}(S_1, \dots, S_n))$.

3.2.3. Z-Transform and Expectation

All results in this subsection are mere specifications of properties which hold for general Markovian arrival processes and were derived in chapter 2. First, a z-transform is given for SMAPs. This is used to determine the expectation of the marginal process $(N(S), J)$ for any set $S \in \sigma(R)$.

Definition 3.8 Let (N, J) be an SMAP with arrival space $(R, \sigma(R))$. According to definition 2.5, the z-transform of (N, J) over the time interval $]s, t]$ is defined as the function $z \rightarrow N(s, t; z)$ with values being the kernels on Φ which are determined by

$$N(s, t; z)(y, B) := \left(\sum_{n=0}^{\infty} P(\text{pr}_S(N_t - N_s) = n, J_t \in B | J_s = y) z^n : S \in \sigma(R) \right)$$

for all $y \in \Phi$, $B \in \sigma(\Phi)$ and $z \in \mathbb{C}$ with $|z| \leq 1$.

Theorem 3.9 *The z -transform of (N, J) over the time interval $]s, t]$ has the form*

$$N(s, t; z) = (N^S(s, t; z) : S \in \sigma(R))$$

with

$$N^S(s, t; z) = \sum_{n=0}^{\infty} \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} \sum_{k=0}^{\infty} D_k(u_1; S) z^k \dots \sum_{k=0}^{\infty} D_k(u_n; S) z^k du_1 \dots du_n$$

for all $S \in \sigma(R)$ and $z \in \mathbb{C}$ with $|z| \leq 1$.

Proof: see theorem 2.9

Remembering definition 2.2, the expectation kernel of an SMAP (which is a kernel on the phase space Φ) can be written as follows:

Theorem 3.10 *The expectation kernel of the marginal process N over the time interval $]s, t]$ is given by*

$$E(N_t - N_s) = \left(\int_s^t P^\Phi(s, u) \sum_{k=1}^{\infty} k \cdot D_k(u; S) P^\Phi(u, t) du : S \in \sigma(R) \right)$$

for all $s < t \in \mathbb{R}_0^+$.

Proof: see theorem 2.10

Example 3.2 As in example 3.1, assume that there is a function $\lambda : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ such that $\Delta(t) = \lambda(t) \cdot \Delta$ for all $t \in \mathbb{R}_0^+$. Further assume that the phase process is homogeneous with stationary distribution π . Then the expectation $E_\pi(N_t - N_s)$ of the marginal process N over the time interval $]s, t]$ given a phase distribution π at time s is

$$E_\pi(N_t - N_s) = \left(\int_s^t \lambda(u) du \cdot \int_\Phi \sum_{k=1}^{\infty} k \cdot D_k(S)(y, \Phi) d\pi(y) : S \in \sigma(R) \right)$$

for all $s < t \in \mathbb{R}_0^+$. This is an obvious consequence of theorem 3.10 noting that $\pi P^\Phi(s, u) = \pi$ and $P^\Phi(u, t)(y, \Phi) = 1$ for all $u \in [s, t]$.

3.2.4. Periodic SMAPs

Now assume that (N, J) is a periodic SMAP. This means that there is a period length $T > 0$ such that the generator of (N, J) satisfies $D(t + T) = D(t)$ for all $t \in \mathbb{R}_0^+$. If further the phase process J has a periodic family of asymptotic distributions (cf. appendix A.3), then some important asymptotic properties can be proven for (N, J) .

3. Spatial Markovian Arrival Processes

Under the condition that the phase process is (periodically) stationary, the expectation has a simpler representation than in the general case. In this subsection, a strong law of large numbers is given which shows that the mean values of the process over long durations tend against this phase-stationary mean rate over a period length. Further, since homogeneous SMAPs are merely special periodic SMAPs, the same results are valid for the homogeneous case, although they assume a simpler form.

Theorem 3.11 *Let (N, J) be a periodic SMAP with arrival space $(R, \sigma(R))$ and period length $T > 0$. Assume that the phase process J is asymptotic with a periodic family $(\pi_t : t \in [0, T[)$ of asymptotic distributions. Then the expectation of the marginal process N after one period length in phase equilibrium (i.e. with initial phase distribution π_0) is given by*

$$E_{\pi_0}(N_T) = \left(\int_0^T \int_{\Phi} d\pi_t(y) \sum_{n=1}^{\infty} n D_n(t; S)(y, \Phi) dt : S \in \sigma(R) \right)$$

Proof: This follows from theorem 3.10 with $s = 0$ and $t = T$ if one acknowledges $\pi_0 P^{\Phi}(0, u) = \pi_u$ for all $u \in [0, T]$ as well as $P^{\Phi}(u, T)(y, \Phi) = 1$ for all $y \in \Phi$ and $u \in [0, T]$.

Theorem 3.12 *Let (N, J) be a periodic SMAP with arrival space $(R, \sigma(R))$ and period length $T > 0$. Assume that the phase process J is asymptotic with a periodic family $(\pi_t : t \in [0, T[)$ of asymptotic distributions. For every set $S \in \sigma(R)$, define the mean arrival rate at S during a period length by*

$$\lambda(S) := \frac{1}{T} \int_0^T \int_{\Phi} d\pi_t(y) \sum_{n=1}^{\infty} n D_n(t; S)(y, \Phi) dt$$

Then for any set $S \in \sigma(R)$, the convergence

$$\frac{N_t(S)}{t} \rightarrow \lambda(S) \text{ as } t \rightarrow \infty$$

holds P -almost surely for all initial distributions.

Proof: This is a corollary to theorem 2.11.

Remark 3.7 As in remark 2.12, the convergence in the above theorem can be proven to hold in terms of a norm $\|\cdot\|$ on the space $\mathbb{R}^{\sigma(R)}$, if $(\mathbb{R}^{\sigma(R)}, \|\cdot\|)$ is a Banach space. In this case, define $\lambda : \sigma(R) \rightarrow \mathbb{R}$ as the global mean rate on $\sigma(R)$. Then the convergence

$$\frac{N_t}{t} \rightarrow \lambda \text{ as } t \rightarrow \infty$$

with respect to the norm $\|\cdot\|$ on $\mathbb{R}^{\sigma(R)}$ holds P -almost surely for all initial distributions.

Example 3.3 Assume that there is a T -periodic function $\lambda : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ such that $\Delta(t) = \lambda(t) \cdot \Delta$ for all $t \in \mathbb{R}_0^+$. Then the mean rate $\lambda(S)$ can be computed as

$$\lambda(S) = \frac{1}{T} \int_0^T \lambda(t) dt \cdot \int_{\Phi} \sum_{n=1}^{\infty} n D_n(S)(y, \Phi) d\pi(y)$$

for all $S \in \sigma(R)$.

Since an homogeneous process with asymptotic distribution π is a periodic process with arbitrary period length $T > 0$ and periodic family $(\pi_t = \pi : t \in [0, T])$ of asymptotic distributions, the same result holds for homogeneous SMAPs, assuming a simpler form:

Theorem 3.13 Let (N, J) be an homogeneous SMAP with arrival space $(R, \sigma(R))$. Assume that the phase process J is asymptotic with stationary distribution π . For every set $S \in \sigma(R)$, define the mean arrival rate at S by

$$\lambda(S) := \int_{\Phi} d\pi(y) \sum_{n=1}^{\infty} n D_n(S)(y, \Phi)$$

Then for any set $S \in \sigma(R)$, the convergence

$$\frac{N_t(S)}{t} \rightarrow \lambda(S) \text{ as } t \rightarrow \infty$$

holds P -almost surely for all initial distributions.

Proof: This is a corollary to theorem 2.12.

Example 3.4 In the case of a BMAP with representation $(D_n : n \in \mathbb{N}_0)$ and m phases, we have the mean rate

$$\lambda = \lambda(R) = \pi \sum_{n=1}^{\infty} n D_n 1_m$$

with π denoting the stationary probability vector of the phase process (i.e. $\pi \sum_{n=0}^{\infty} D_n = 0$) and 1_m denoting the m -dimensional column vector with all entries being 1.

3.3. Examples

Throughout this section, all transition rates $D_{k_1, \dots, k_n}(S_1, \dots, S_n)(y, A)$ which are not defined explicitly, shall be assumed as zero.

3. Spatial Markovian Arrival Processes

Example 3.5 If the arrival space $R = \{r\}$ of an SMAP (N, J) is trivial, i.e. consisting of only one element, then (N, J) shall be called a **generalized BMAP** or shortly **gBMAP**. A generalized BMAP which is periodic with some period length $T > 0$ shall be called **periodic BMAP** or shortly **pBMAP**. The most important difference to the BMAP concept introduced by Lucantoni [60] is that generalized BMAPs provide for inhomogeneous transition rates.

By specification of the arrival space $R = \{r\}$, chapter 4 includes an analysis of the pBMAP/M/c queue with periodic service rates and chapter 5 contains an examination of the gBMAP/G/ ∞ queue as a special case.

Example 3.6 Let the arrival space R be separable and choose any upper bound $M \in \mathbb{R}^+$ for the arrival intensities at any point in R . Then the space $C(R)$ of continuous functions from R into the compact interval $[0, M]$ is a compact and separable metric space. This follows from standard topological results (cf. Herrlich [48], pp.206,118,117) and the fact that any continuous function is determined by its values on a countable dense subset of the separable space R . Define the phase space to be $\Phi := C(R)$ and let $\sigma(\Phi)$ denote the Borel σ -algebra on Φ . Since every function $f \in \Phi$ is continuous and hence uniquely determined by its values on a countable dense subset of R , it follows that $\{f\} \in \sigma(\Phi)$ for every $f \in \Phi$ (cf. Neveu [74], p.103). Phases out of this space can serve as intensity functions for arrivals distributed by inhomogeneous Poisson point fields (see Kingman [55]).

A **spatial Markov modulated Poisson process** (or shortly **spatial MMPP**) can be defined as follows: Let D denote the generator of an homogeneous Markov jump process with state space Φ as described above and Π_f an inhomogenous Poisson process on R (see Kingman [55]) with intensity $\int_S f(x)d\mu(x)$ for $S \in \sigma(R)$. Further, let $\gamma_f > 0$ for all $f \in \Phi$ such that $\sup_{f \in \Phi} \gamma_f < \infty$. The transition rates are defined as

$$D_{0,\dots,0}(S_1, \dots, S_n)(f, A) := D(f, A) - \gamma_f(1 - \Pi_f(S_1, \dots, S_n)(0, \dots, 0))$$

for $A \in \sigma(\Phi)$ and $f \in A$,

$$D_{0,\dots,0}(S_1, \dots, S_n)(f, A) := D(f, A)$$

for $A \in \sigma(\Phi)$ and $f \notin A$ and

$$D_{k_1,\dots,k_n}(S_1, \dots, S_n)(f, A) := \gamma_f \cdot \Pi_f(S_1, \dots, S_n)(k_1, \dots, k_n)$$

for $\sum_{i=1}^n k_i \geq 1$, $A \in \sigma(\Phi)$ and $f \in A$.

Obviously, Π_f does not need to be a Poisson process on R and could be chosen as any other modulated random point field instead.

Example 3.7 If the phase space $\Phi = \{\varphi\}$ of an SMAP (N, J) is trivial, i.e. consisting of only one element, then (N, J) shall be called **general Poisson process with spatial arrivals**. For these special SMAPs, the kernels $D_{k_1,\dots,k_n}(S_1, \dots, S_n)$ on Φ become real numbers and all computations are much more tractable. If even more the arrival space $R = \{r\}$ is trivial and (N, J) is homogeneous in time, then (N, J) is a Poisson process with independent and identically distributed batch arrivals.

Example 3.8 A more complicated, but still naturally appearing example is the following **spatially ordered superposition of independent BMAPs**:

Let $\mathcal{A} = \{A_k : k \in \mathbb{N}\}$ denote a denumerable partition of R with $A_k \in \sigma(R)$ for all $k \in \mathbb{N}$. For every A_k , define a BMAP by its generating sequence $\{D_n(A_k) : n \in \mathbb{N}_0\}$. Denote by $\sigma(\mathcal{A})$ the smallest σ -algebra on the set system \mathcal{A} . Assume that $\sum_{k=1}^{\infty} \|D_0(A_k)\| < \infty$ with $\|D\|$ denoting the maximum norm of a matrix D .

The spatially ordered superposition of the above BMAPs yields an SMAP. Its phases $\varphi \in \Phi$ can be described by functions $\varphi : \mathbb{N} \rightarrow \mathbb{N}$, with $\varphi(k)$ denoting the phase of the BMAP over A_k .

For $S \in \sigma(\mathcal{A})$ and $\varphi \in \Phi$, the one-dimensional transition rates are given as follows:

$$D_n(S)(\varphi, \{\psi\}) = D_{n;\varphi(k),\psi(k)}(A_k)$$

for $n \in \mathbb{N}_0$, $A_k \subset S$, $\psi(k) \neq \varphi(k)$ and $\psi(m) = \varphi(m)$ for all $m \neq k$,

$$D_0(S)(\varphi, \{\psi\}) = \sum_{n=0}^{\infty} D_{n;\varphi(k),\psi(k)}(A_k)$$

for $A_k \subset R \setminus S$, $\psi(k) \neq \varphi(k)$ and $\psi(m) = \varphi(m)$ for all $m \neq k$, and

$$D_n(S)(\varphi, \{\varphi\}) = \sum_{A_k \subset S} D_{n;\varphi(k),\varphi(k)}(A_k)$$

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

The diagonal elements are

$$D_0(S)(\varphi, \{\varphi\}) = \sum_{k=1}^{\infty} D_{0;\varphi(k),\varphi(k)}(A_k) + \sum_{A_k \subset R \setminus S} \sum_{n=1}^{\infty} D_{n;\varphi(k),\varphi(k)}(A_k)$$

For $A \in \sigma(\Phi)$ and $\varphi \in \Phi$, the set of functions ψ , for which $\{k \in \mathbb{N} : \psi(k) \neq \varphi(k)\}$ has at most one element, is countable. Since for all other functions ψ the transition rates $D_n(S)(\varphi, \{\psi\})$ are zero, the sums in

$$D_n(S)(\varphi, \Psi) = \sum_{\psi \in \Psi} D_n(S)(\varphi, \{\psi\})$$

are well-defined for all $n \in \mathbb{N}_0$.

The transition rates on finite families of disjoint sets can be expressed in terms of the one-dimensional transition rates as follows:

Be $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$ disjoint. Since for a superposition, there can be at most one batch arrival in one set during an infinitesimal time interval, we can define

$$D_{k_1, \dots, k_n}(S_1, \dots, S_n)(\varphi, \{\psi\}) := D_{k_i}(S_i)(\varphi, \{\psi\}),$$

3. Spatial Markovian Arrival Processes

if $k_j = 0$ for all $j \neq i$, $A_k \subset S_i$, $\psi(k) \neq \varphi(k)$ and $\psi(m) = \varphi(m)$ for all $m \neq k$,

$$D_{k_1, \dots, k_n}(S_1, \dots, S_n)(\varphi, \{\psi\}) := D_0 \left(\bigcup_{i=1}^n S_i \right) (\varphi, \{\psi\}),$$

if $\sum_{i=1}^n k_i = 0$, $A_k \subset R \setminus \bigcup_{i=1}^n S_i$, $\psi(k) \neq \varphi(k)$ and $\psi(m) = \varphi(m)$ for all $m \neq k$,

$$D_{k_1, \dots, k_n}(S_1, \dots, S_n)(\varphi, \{\varphi\}) := D_{k_i}(S_i)(\varphi, \{\varphi\}),$$

if $k_i \in \mathbb{N}$ and $k_j = 0$ for all $j \neq i$, and

$$D_{k_1, \dots, k_n}(S_1, \dots, S_n)(\varphi, \{\varphi\}) := D_0 \left(\bigcup_{i=1}^n S_i \right) (\varphi, \{\varphi\}),$$

if $\sum_{i=1}^n k_i = 0$.

Again, the sums in

$$D_{k_1, \dots, k_n}(S_1, \dots, S_n)(\varphi, \Psi) = \sum_{\psi \in \Psi} D_{k_1, \dots, k_n}(S_1, \dots, S_n)(\varphi, \{\psi\})$$

are countable and thus well-defined.

If the distribution of the spatial arrivals is neglected, then, analogously to the case of a superposition of finitely many BMAPs (see Baum [14]), the generating sequence $\Delta(R)$ of the superposition over the whole space R is the Kronecker sum of the generating sequences of its component BMAPs.

4. Markovian Spatial Queues

In most textbooks on queueing theory, Markovian queues are the first ones that are analyzed. This clearly is a result of the easy structure of Markovian queues, which can be analyzed as a Markov jump process in a straightforward manner. For this reason, in the present thesis Markovian queues will be examined first. However, this will be done on a more general level, covering the important case of periodic Markovian queues.

A typical property of communication traffic is the dependence of its arrival rates on time. This aspect incites the use of time-inhomogeneous processes and queues for modelling communication networks. Typically, a periodic dependence of the arrival rates and/or the service time distribution can be assumed with period lengths of a day or a week.

While queues with periodic input naturally reflect the time-dependent amount of traffic that arrives in communication networks, the analysis of queues with inhomogeneous arrival rates is far less developed than the one for homogeneous queues. Some of the existing results in the literature are given in Asmussen and Thorisson [5], Bambos and Walrand [10], Falin [38], Harrison and Lemoine [46], Hasofer [47], Heyman and Whitt [49], Lemoine [57, 56], Rolski [82, 83], and Willie [94].

In this chapter, two Markovian queues, which have not been analyzed before, will be examined. The first one is the spatial $SMAP/M_t/c/c$ queue with a spatial Markovian arrival process, a general (i.e. possibly inhomogeneous) Poisson service process, c servers and an equal system capacity. The last condition prevents an arriving user, who finds all c servers busy, from entering the system, such that this job is lost. This is a natural assumption for the application field of mobile communication systems, since a telephone customer who finds a busy line will not wait but choose another net provider. For this queue, a loss formula will be derived. If one wishes to approximate a non-Poisson service process, the same method of analysis can be applied to the $SMAP/PH_t/c/c$ queue with phase-type service time distribution PH_t that may have time-dependent rates (see remark 2.7).

The second queue to be examined in this chapter is the periodic $BMAP/M_t/c$ queue with periodic service rates. This is a special case of the $SMAP/M_t/c$ queue with inhomogeneous service rates. Whereas the $SMAP/M_t/c/c$ queue with finite phase space is a finite state Markov jump process and hence has an asymptotic distribution, a stability condition must be derived for the periodic $BMAP/M_t/c$ queue. The same method of analysis applies to the periodic $BMAP/PH_t/c$ queue if one wishes to further approximate a non-Poisson service process.

4.1. The $SMAP/M_t/c/c$ Queue

The queue examined in this section is defined as follows. The arrival process is an SMAP. The number c of servers is finite and equals the system capacity for users. This means that whenever an arriving user finds all servers busy, he cannot enter the system and is rejected. Hence there are no waiting users in the queue. Every server is equal, and the service time distribution function B_s for a user arriving at time $s \in \mathbb{R}_0^+$ is defined by

$$B_s(t - s) := 1 - e^{-\int_s^t \mu_u du}$$

for all $t > s$, with $\mu_u \in \mathbb{R}_0^+$ for all $u \in [s, t]$. This means that the service process without idle periods would be an inhomogeneous Poisson process with rates $(\mu_t : t \in \mathbb{R}_0^+)$.

The $SMAP/M_t/c/c$ queue is a natural model for a cell of an FDMA or a TDMA mobile communication network (cf. Gibson [42] or Rappaport [81]). For such a cell, which remains of constant size over time (unlike a CDMA cell, cf. Yang [96] or Viterbi [93]), the arrival space R of the queue would be the landscape covered by the cell. The number c of servers would be the number of available channels, i.e. the ratio of the total band width to a single frequency lot assigned to every user (in the FDMA case). The assumption that no user is waiting appears natural, as remarked above. Furthermore, usually the arrival process will be periodic with the period length being a day or a week.

In order to derive the simplest possible form of a loss formula (see section 4.1.3), the rejection policy of the communication network to be modelled shall be defined as follows. A rejection shall be conceived to occur only in the event that the system capacity is totally filled, i.e. that there are c users in the system. Thus, if there are $k < c$ users in the system and a batch arrival of size h with $k + h > c$ occurs, then this will not be rejected although it cannot be totally served by the system. In this case, the queue will be filled to its full capacity c . This can be interpreted as reflecting the interest of the network provider to partially serve an incoming user instead of rejecting him, even if the total amount of service that the user requires would exceed the system capacity. It is possible to derive transient and asymptotic distributions as well as a loss formula for any other rejection policy using the same methods, but this would result in more complicated formulae.

Let (N, J) denote the SMAP of the queue with arrival space $(R, \sigma(R))$ and

$$(\Delta(t; S_1, \dots, S_n) : t \in \mathbb{R}_0^+, n \in \mathbb{N}, S_1, \dots, S_n \in \sigma(R))$$

denoting the generating sequences of (N, J) (see definition 3.4). Further, let $(\mu_t : t \in \mathbb{R}_0^+)$ denote the time-dependent service rate, which shall be equal for all servers.

For ease of notation, the queue distribution shall be derived only for one-dimensional marginal processes $Q(S)$, i.e. only one subset $S \in \sigma(R)$ will be described. For higher-dimensional marginal distributions on (S_1, \dots, S_n) with $n > 1$, the method of analysis is completely analogous.

Denote the joint queueing process in the subsets R and S by $Q(R, S) = (Q_t(R, S) : t \in \mathbb{R}_0^+)$. The infinitesimal generator $G(t; R, S)$ of $Q(R, S)$ can be written in a block matrix with entries being the kernels

$$G_{(h,k),(l,n)}(t; R, S) = \begin{cases} k\mu_t \cdot I & \text{for } h = l + 1 \leq c, k = n + 1 \leq l + 1 \\ (h - k)\mu_t \cdot I & \text{for } h = l + 1 \leq c, k = n \leq l \\ D_{0,0}(t; R, S) - h\mu_t \cdot I & \text{for } h = l < c, k = n \leq l \\ D_{0,0}(t; R, S) - h\mu_t \cdot I \\ \quad + \sum_{i=1}^{\infty} D_i(t; R) & \text{for } h = l = c, k = n \leq l \\ D_{l-h,n-k}(t; R, S) & \text{for } h < l < c, k \leq n \leq l \\ \sum_{i=c-h}^{\infty} D_{i,n-k}(t; R, S) & \text{for } h < l = c, k \leq n \leq l \\ 0 & \text{otherwise} \end{cases}$$

for all $h, k, l, n \in \{0, \dots, c\}$, with 0 and I denoting the zero and identity kernel, respectively.

Remember that the $D_{l,n}(t; R, S)$ are kernels on the phase space Φ of the SMAP (N, J) with $D_{l,n}(t; R, S)(y, A)$ denoting the infinitesimal transition rate at time t in phase y to change to some phase $z \in A$ and observe n arrivals in S as well as l arrivals in R . Thus the above definition of the generator means that a phase transition may occur even if the arrival occurring at the same time is rejected.

4.1.1. Transient Distribution

As a Markov jump process, the queue allows a description of its transient distribution in terms of the transition probabilities. First, this shall be expressed by formula A.2. Then, the obtained general form will be simplified for special cases.

The multiplication of two block matrices K and L with kernel entries is defined by

$$(KL)_{(h,k),(l,n)}(y, A) := \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \int_{\Phi} K_{(h,k),(i,j)}(y, dz) L_{(i,j),(l,n)}(z, A)$$

which for the above generators can be simplified to

$$(KL)_{(h,k),(l,n)}(y, A) = \sum_{i=0}^c \sum_{j=0}^i \int_{\Phi} K_{(h,k),(i,j)}(y, dz) L_{(i,j),(l,n)}(z, A)$$

for all $h, k, l, n \in \{0, \dots, c\}$, $y \in \Phi$ and $A \in \sigma(\Phi)$.

Define $P_{(h,k),(l,n)}(s, t)(y, A)$ as the transition probability of having (l, n) users in (R, S) and being in some phase $z \in A$ at time $t > s$ under the condition of having (h, k) users in (R, S) and being in phase y at time s . Further define $P_{(h,k),(l,n)}(s, t)$ as the sub-Markov kernel on the phase space Φ which results from the probabilities $P_{(h,k),(l,n)}(s, t)(y, A)$ and $P(s, t) = (P_{(h,k),(l,n)}(s, t))$ as the block matrix with kernel entries $P_{(h,k),(l,n)}(s, t)$.

4. Markovian Spatial Queues

Theorem 4.1 *The transition probability kernel $P(s, t; R, S)$ of the queue $Q(R, S)$ from time s to time t can be written as*

$$P(s, t; R, S) = \sum_{k=0}^{\infty} \underbrace{\int_s^t \int_s^{u_k} \dots \int_s^{u_2}}_{k \text{ integrals}} G(u_1; R, S) \dots G(u_k; R, S) du_1 \dots du_k \quad (4.1)$$

defining the summand for $k = 0$ as the identity matrix Id with entries $\delta_{kn} \cdot I$.

Proof: The queue process $Q(R, S)$ is a Markov jump process. Hence, formula A.2 applies for the transition probabilities.

☺

Now, fix a time $t \in \mathbb{R}_0^+$ at which the queue is to be observed. Denote the initial distribution of the queue process at time $s := 0$ by ν . Let $Q_{(l,n)}(t; R, S)(y, A)$ denote the probability that the queue has (l, n) users in (R, S) and is in some phase $z \in A$ at time t under the condition of having $(0, 0)$ users in (R, S) and being in phase y at time 0. Further define $Q_{(l,n)}(t; R, S)$ as the sub-Markov kernel on the phase space Φ which results from the transient probabilities $Q_{(l,n)}(t; R, S)(y, A)$ at time $t \in \mathbb{R}^+$ and $Q_t(R, S) = (Q_{(l,n)}(t; R, S))$ as the block matrix with kernel entries $Q_{(l,n)}(t; R, S)$. Then the transient distribution at time t is determined by

$$\begin{aligned} Q_t(R, S) &= \int d\nu P(0, t; R, S) \\ &= \int d\nu \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_k} \dots \int_0^{u_2}}_{k \text{ integrals}} G(u_1; R, S) \dots G(u_k; R, S) du_1 \dots du_k \end{aligned}$$

defining the summand for $k = 0$ as the identity matrix Id . This expression assumes a simpler form for the following special cases.

The simplest form can be arrived at in the homogeneous case. Then the transition probabilities assume the usual exponential form which holds for homogeneous Markov jump processes.

Theorem 4.2 *For homogeneous arrival and service rates, the generator $G(R, S) := G(t; R, S)$ is constant in t and the transient distribution at time $t \in \mathbb{R}^+$ is determined by*

$$Q_t(R, S) = \int d\nu \sum_{k=0}^{\infty} \frac{t^k}{k!} G(R, S)^k = \int d\nu e^{G(R, S) \cdot t}$$

with $G(R, S)^k$ denoting the k -th power of the kernel matrix $G(R, S)$.

Proof: This is immediate from theorem A.8.

Remark 4.1 According to theorem A.9, it is possible to approximate the general case by a sequence of piecewise homogeneous processes. For this approximation, the transient distribution can be computed as a product of exponential forms.

Another case of simplification is that of quasi-commutable generators (see definition A.8). Here, an exponential form is assumed, too, although with a more complicated exponent.

Theorem 4.3 *If the generators $(G(t; R, S) : t \in \mathbb{R}_0^+)$ are quasi-commutable, which means that the equation*

$$\frac{d}{dt} \left(\int_0^t G(u; R, S) du \right)^k = k \cdot \left(\int_0^t G(u; R, S) du \right)^{k-1} G(t; R, S)$$

holds for all $k \in \mathbb{N}$ and $t \in \mathbb{R}_0^+$, the transient distribution of $Q(R, S)$ with initial distribution ν takes the form

$$Q_t(R, S) = \int d\nu \sum_{k=0}^{\infty} \frac{\left(\int_0^t G(u; R, S) du \right)^k}{k!} = \int d\nu e^{\int_0^t G(u; R, S) du}$$

Proof: see theorem A.22

Remark 4.2 A simple form of quasi-commutability is given if the generators have the form

$$G(t; R, S) = \lambda(t) \cdot G(R, S)$$

for all $t \in \mathbb{R}_0^+$, with $\lambda : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ being a non-negative function in time and $G(R, S)$ denoting a generator which is constant in time (cf. remark A.9). This would for example be the case if the arrival process satisfied $\Delta(t) = \lambda(t) \cdot \Delta$, with Δ being constant in time, and the service rates were constant.

Most important for applications in modelling mobile communication networks is the case of periodic arrival and service rates. In this case, a computation of the transient distribution can be simplified as follows. For ease of notation, write $P(s, t) := P(s, t; R, S)$ for all $s < t \in \mathbb{R}_0^+$. The periodicity of the generator yields

$$P(0, nT) = P(0, (n-1)T)P((n-1)T, nT) = P(0, (n-1)T)P(0, T) = P(0, T)^n$$

Let ν denote the initial distribution of the queue process Q . Define

$$\lfloor t/T \rfloor := \max\{n \in \mathbb{N}_0 : nT \leq t\}$$

as the number of period lengths that have passed until time $t \in \mathbb{R}^+$. Now the transient distribution of Q with initial distribution ν is given by

$$Q_t^\nu = \int d\nu P(0, \lfloor t/T \rfloor T) P(\lfloor t/T \rfloor T, t) = \int d\nu P(0, T)^{\lfloor t/T \rfloor} P(0, t - \lfloor t/T \rfloor T)$$

This expression allows a computation of the transient distribution at any time $t \in \mathbb{R}^+$ without needing to integrate over ranges larger than the period T . For computing the remaining terms

4. Markovian Spatial Queues

$P(0, s)$ with $s \leq T$, one can use the following iteration as given in Bellman [19], p.168: Starting with $I_0(u) := Id$ for all $u \leq s$, the iteration

$$I_{n+1}(u) := \int_0^u I_n(v)G(v; R, S)dv + Id$$

leads to the limit

$$P(0, s) = \lim_{n \rightarrow \infty} I_n(s)$$

for all $s \leq T$.

Example 4.1 Assume that the parameters are given as follows: Let Δ denote the generating sequence of an homogeneous SMAP and choose service rates $\mu_1, \mu_2 \in \mathbb{R}^+$. Further, let $T > 0$ denote a period length and set

$$\Delta(t) = \left(1 + \sin\left(\frac{2\pi}{T}t\right)\right) \cdot \Delta$$

as well as

$$\mu_t = \begin{cases} \mu_1 & \text{for } t < T/2 \\ \mu_2 & \text{for } t \geq T/2 \end{cases}$$

for all $t \in [0, T[$. Then the generators are periodic with period T and piecewise quasi-commutable (on $[0, T/2[$ and on $[T/2, T[$). Denote $G_1 := G(0; R, S)$ as well as $G_2 := G(T/2; R, S)$. According to example A.4, we then obtain

$$P(0, s) = \exp\left(\left(s + \frac{T}{2\pi}\left(1 - \cos\left(\frac{2\pi}{T}s\right)\right)\right) \cdot G_1\right)$$

for $s < T/2$ and

$$P(0, s) = \exp\left(\left(\frac{T}{2} + \frac{T}{\pi}\right) G_1\right) \exp\left(\left(s - \frac{T}{2} - \frac{T}{2\pi}\left(1 + \cos\left(\frac{2\pi}{T}s\right)\right)\right) \cdot G_2\right)$$

for $s \geq T/2$.

4.1.2. Asymptotic Distributions

In the case of a periodic $SMAP/M_t/c/c$ queue with finite phase space, a periodic family of asymptotic distributions can be given. This is an immediate consequence of the results for periodic Markov jump processes in section A.3.

Let Q denote an $SMAP/M_t/c/c$ queue with finite phase space $\Phi = \{1, \dots, m\}$. Further, let $S \in \sigma(R)$ denote any subset of the arrival space and $Q(R, S)$ the marginal process of Q on the subsets R and S . Assume that $Q(R, S)$ is periodic with period length $T > 0$.

Let $Y(R, S) = (Y_n(R, S) : n \in \mathbb{N}_0)$ be the homogeneous discrete time Markov chain with transition kernel

$$P(0, T; R, S) = \sum_{j=0}^{\infty} \int_0^T \int_0^{u_j} \dots \int_0^{u_2} G(u_1; R, S) \dots G(u_j; R, S) du_1 \dots du_j$$

Then the main result is

Theorem 4.4 *The Markov chain $Y(R, S)$ has an asymptotic distribution $q(R, S)$, and the marginal queue process $Q(R, S)$ has a periodic family $(q_s(R, S) : s \in [0, T[)$ of asymptotic distributions (see definition A.13) assuming the form*

$$q_s(R, S) = q(R, S)P(0, s; R, S)$$

for all $s \in [0, T[$. The distribution $q(R, S)$ is determined by the equilibrium equation

$$q(R, S) = q(R, S)P(0, T; R, S)$$

Proof: Since the phase space Φ is finite and the queue has finite capacity, the state space of the queue process is finite, too. It follows that the Markov chain $Y(R, S)$ has finite state space and hence an asymptotic distribution $q(R, S)$. The rest follows from theorem A.28 and remark A.11.

☺

The homogeneous SMAP/ $M/c/c$ queue can be seen as a special case of the periodic SMAP/ $M_t/c/c$ queue with arbitrary period length $T > 0$ and generators $G(R, S)$ which are constant in time. This leads to

Theorem 4.5 *Let Q be an homogeneous SMAP/ $M/c/c$ queue with finite phase space $\Phi = \{1, \dots, m\}$ and generator $G(R, S)$ for the marginal process $Q(R, S)$. Then $Q(R, S)$ has an asymptotic distribution $q(R, S)$ with*

$$q(R, S)G(R, S) = 0$$

Proof: This follows from the above theorem 4.4 and the form

$$P(0, t; R, S) = e^{G(R, S) \cdot t} = \sum_{j=0}^{\infty} \frac{t^j}{j!} G(R, S)^j$$

for the transition probability kernel of $Q(R, S)$ (see theorem A.8). The homogeneous Markov chain $Y(R, S)$ defined above has an asymptotic distribution $q(R, S)$, since the state space is finite. This satisfies $q(R, S)G(R, S) = 0$ and because of the above exponential form $q(R, S)P(0, s; R, S) = q(R, S)$ for all $s \in [0, T[$.

☺

4. Markovian Spatial Queues

Remark 4.3 In the homogeneous case, the generator matrix

$$G(R, S) = \left(G_{(h,k),(l,n)}(R, S) \right)_{0 \leq k \leq h \leq c, 0 \leq n \leq l \leq c}$$

can be arranged according to the indices (k, n) . This results in an $M/G/1$ -type matrix and hence the distribution $q(R, S)$ can be found by the method introduced in Ramaswami [80] and generalized by Hofmann [50].

4.1.3. A Loss Formula

The $SMAP/M_t/c/c$ queue is a **loss system**. This means that arriving users who find all servers busy are rejected and since they do not retry to enter the queue, they can be regarded as lost. One of the most important performance measures of loss systems is the loss probability, which is defined as the long time fraction of lost, i.e. rejected users to all users having arrived. For the application field of mobile communication networks, the meaning of this probability as a criterion of quality is immediate.

For $M/G/c/c$ loss systems, the loss probability is the asymptotic probability p_c of having c users in the system, which is the same event as all servers being busy. This result is independent of the service time distribution and was first exhibited by Sevast'yanov [86] in 1957. It follows easily from the PASTA property for queues with Poisson arrival processes (cf. Wolff [95], pp.271-273). In this subsection, the loss probability will be derived using the law of large numbers for periodic SMAPs (theorem 2.11). It turns out that an analogous result holds, although the PASTA property does not hold for the $SMAP/M_t/c/c$ queue (for a more general treatment of the PASTA problem see Melamed, Yao [65]).

Definition 4.1 Let Q denote a loss system. Denote the number of users, that have arrived at Q until time $t \in \mathbb{R}_0^+$, by N_t and the number of rejected users until time $t \in \mathbb{R}_0^+$ by L_t . Assume that Q has an asymptotic distribution. Then the **loss probability** p_l of Q is defined as the fraction

$$p_l := \lim_{t \rightarrow \infty} \frac{L_t}{N_t}$$

of the number of rejected users over the number of all users having arrived in the long run.

Now let Q denote the periodic $SMAP/M_t/c/c$ queue with arrival space $(R, \sigma(R))$, finite phase space $\Phi = \{1, \dots, m\}$ and period length $T > 0$. As usual, notational convenience is the reason for analyzing only the one-dimensional marginal processes on any measurable set $S \in \sigma(R)$. Denote the SMAP arrival process on the subset S by $(N(S), J)$ and define it by its generating sequence $(D_n(S) : l, n \in \mathbb{N}_0)$. Then $D := \sum_{n=0}^{\infty} D_n(S)$ is the generator of the phase process J . This has a periodic family $(\pi_s : s \in [0, T])$ of asymptotic distributions, since the phase space Φ is finite. Furthermore, the finiteness of Φ allows to represent the measures π_s as row vectors in \mathbb{R}^m and the kernels $D_n(S)$ and D as $m \times m$ matrices with real-valued

entries. Define the column vector $1_m \in \mathbb{R}^m$ as the one with all entries being 1. Finally, assume that the asymptotic mean arrival rate during one period length

$$\lambda(S) := \frac{1}{T} \int_0^T \pi_t \sum_{n=1}^{\infty} n D_n(t; S) 1_m dt < \infty$$

is finite. Then the main result is

Theorem 4.6 *Let Q denote a periodic SMAP/ $M_t/c/c$ loss system with finite phase space $\Phi = \{1, \dots, m\}$, arrival space $(R, \sigma(R))$ and period length $T > 0$. Let the SMAP of Q have a generating sequence $(D_n(S) : n \in \mathbb{N}_0)$ on the subset $S \in \sigma(R)$. Assume that $Q(R)$ has a periodic family $(q_s : s \in [0, T])$ of asymptotic distributions and define $q_t(c) \in \mathbb{R}^m$ as the row vector with i th entry $q_t(c, i)$. Then the loss probability of Q on the subset S is*

$$p_l = \frac{\int_0^T q_t(c) \sum_{n=1}^{\infty} n D_n(t; S) 1_m dt}{\int_0^T \pi_t \sum_{n=1}^{\infty} n D_n(t; S) 1_m dt}$$

for all $S \in \sigma(R)$.

Proof: For ease of notation, we shall write $N_t := N_t(S)$, $L_t := L_t(S)$ and $D_n(t) := D_n(t; S)$ in this proof. By definition, the arrival process $N(S)$ in S is a one-dimensional Markovian arrival process. Thus, the strong law of large numbers (theorem 2.11) applies and yields

$$\lim_{t \rightarrow \infty} \frac{N_t}{t} = \frac{1}{T} \int_0^T \pi_t \sum_{n=1}^{\infty} n D_n(t) 1_m dt \quad (4.2)$$

almost surely.

The process $L(S) = (L_t : t \in \mathbb{R}_0^+)$ which counts the number of rejected users can be described by a thinned Markovian arrival process. The infinitesimal transition rates of $L(S)$ are given by

$$\begin{aligned} R_n(t; i, j) &= \lim_{\Delta t \rightarrow 0} \frac{P(Q_t(R) = (c, i), N_{t+\Delta t} - N_t = n, J_{t+\Delta t} = j | J_t = i)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(N_{t+\Delta t} - N_t = n, J_{t+\Delta t} = j | Q_t(R) = (c, i), J_t = i)}{\Delta t} \\ &\quad \cdot P(Q_t(R) = (c, i) | J_t = i) \\ &= \lim_{\Delta t \rightarrow 0} \frac{P(N_{t+\Delta t} - N_t = n, J_{t+\Delta t} = j | J_t = i)}{\Delta t} \cdot P(Q_t(R) = (c, i) | J_t = i) \\ &= D_n(t; i, j) \cdot P(Q_t(R) = (c, i) | J_t = i) \end{aligned}$$

for every $n \in \mathbb{N}$, since the arrival process $(N(S), J)$ only depends on the phase but not on the number of users in R . Furthermore, we have

$$R_0(t; i, j) = D_0(t; i, j) + \sum_{n=1}^{\infty} D_n(t; i, j) \cdot (1 - P(Q_t(R) = (c, i) | J_t = i))$$

4. Markovian Spatial Queues

Writing $t = kT + s$ and letting $k \rightarrow \infty$, the asymptotic rates are

$$\lim_{k \rightarrow \infty} R_n(kT + s; i, j) = D_n(s; i, j) \cdot \frac{q_s(c, i)}{\pi_s(i)}$$

for all $s \in [0, T[$, $n \in \mathbb{N}$ and $i, j \in \Phi$ and

$$\lim_{k \rightarrow \infty} R_0(kT + s; i, j) = D_0(s; i, j) + \sum_{n=1}^{\infty} D_n(s; i, j) \cdot \left(1 - \frac{q_s(c, i)}{\pi_s(i)}\right)$$

for all $s \in [0, T[$ and $i, j \in \Phi$. Obviously, the generator of the phase process of $R(S)$ equals

$$\sum_{n=0}^{\infty} R_n(t; i, j) = D(t; i, j)$$

which is the generator of J . Hence, by theorem 2.11, the strong law of large numbers assures the convergence

$$\lim_{t \rightarrow \infty} \frac{L_t}{t} = \frac{1}{T} \int_0^T \pi_t \sum_{n=1}^{\infty} n R_n(t) 1_m dt = \frac{1}{T} \int_0^T q_t(c) \sum_{n=1}^{\infty} n D_n(t) 1_m dt \quad (4.3)$$

almost surely.

Combining the results 4.2 and 4.3, the loss probability is determined by

$$p_l = \lim_{t \rightarrow \infty} \frac{L_t}{t} \cdot \frac{t}{N_t} = \frac{\int_0^T q_t(c) \sum_{n=1}^{\infty} n D_n(t) 1_m dt}{\int_0^T \pi_t \sum_{n=1}^{\infty} n D_n(t) 1_m dt}$$

☺

As usual, the case of an homogeneous queue can be regarded as the special periodic queue with arbitrary period length $T > 0$ and a generator which is constant in time. Thus, the result for homogeneous queues follows immediately from the above theorem.

Theorem 4.7 *Let Q denote an homogeneous SMAP/M/c/c loss system with arrival space $(R, \sigma(R))$. Let the SMAP of Q have a generating sequence $(D_n(S) : n \in \mathbb{N}_0)$ on the subset $S \in \sigma(R)$. Assume that $Q(R)$ has an asymptotic distribution q and define $q(c) \in \mathbb{R}^m$ as the row vector with i th entry $q(c, i)$. Then the loss probability of Q on the subset S is*

$$p_l = \frac{q(c) \sum_{n=1}^{\infty} n D_n(S) 1_m}{\pi \sum_{n=1}^{\infty} n D_n(S) 1_m}$$

for all $S \in \sigma(R)$.

Proof: This merely is the form of p_l (derived in the above theorem 4.6) for the special case of an homogeneous queue.

4.2. The Periodic $BMAP/M_t/c$ Queue

Let $Q = (Q_t : t \in \mathbb{R}_0^+)$ denote a periodic $BMAP/M_t/c$ queue with periodic BMAP input having finite phase space $\{1, \dots, m\}$ (see example 3.5), exponential service time distributions with periodic service rates, and c servers. Arrival process and service rates shall have the same period T . Define the arrival process by its transition rate matrices $(D_n(t) : n \in \mathbb{N}_0, t \in \mathbb{R}_0^+)$ having dimension m and period T . Further, let $(\mu_t : t \in \mathbb{R}_0^+)$ denote the time-dependent service rates, which shall be equal for all servers. Periodicity of the service rates means $\mu_{s+T} = \mu_s$ for all $s \in [0, T[$.

4.2.1. Transient Distributions

As a Markovian queue, the system can be analyzed like an (inhomogeneous) Markov jump process. The infinitesimal generator $G(t)$ of the queue Q can be written as an $\mathbb{N}_0 \times \mathbb{N}_0$ block matrix with entries

$$G_{kn}(t) = \begin{cases} 0 & \text{for } k > n + 1 \\ k\mu_t \cdot I & \text{for } k = n + 1 \leq c \\ c\mu_t \cdot I & \text{for } k = n + 1 > c \\ D_0(t) - k\mu_t \cdot I & \text{for } k = n \leq c \\ D_0(t) - c\mu_t \cdot I & \text{for } k = n > c \\ D_{n-k}(t) & \text{for } k < n \end{cases}$$

for $k, n \in \mathbb{N}_0$, with 0 and I denoting the zero and identity matrix, respectively. Note that the entries are $m \times m$ matrices. Denote the (i, j) th entry of the matrix $G_{kn}(t)$ by $G_{kn}(t)(i, j)$.

The multiplication of two $\mathbb{N}_0 \times \mathbb{N}_0$ block matrices A and B is defined by

$$(AB)_{kn}(i, j) := \sum_{l=0}^{\infty} \sum_{h=1}^m A_{kl}(i, h) B_{ln}(h, j)$$

for every $k, n \in \mathbb{N}_0$ and $i, j \in \{1, \dots, m\}$.

Define $P_{kn}(s, t)(i, j)$ as the probability of having $n \in \mathbb{N}_0$ users in the queue and being in phase j at time $t > s$ under the condition of having $k \in \mathbb{N}_0$ users in the queue and being in phase i at time s . Further define $P_{kn}(s, t)$ as the $m \times m$ matrix with entries $P_{kn}(s, t)(i, j)$ and $P(s, t) = (P_{kn}(s, t))_{k, n \in \mathbb{N}_0}$ as the $\mathbb{N}_0 \times \mathbb{N}_0$ block matrix with entries $P_{kn}(s, t)$.

According to theorem A.6 and formula A.2, the transition probabilities of the queue can be written as

$$P(s, t) = \sum_{k=0}^{\infty} \underbrace{\int_s^t \int_s^{u_k} \dots \int_s^{u_2}}_{k \text{ integrals}} G(u_1) \dots G(u_k) du_1 \dots du_k$$

4. Markovian Spatial Queues

Let ν denote the initial distribution of the queue process Q . The transient distribution of Q is given as in section A.3 by

$$Q_t = \int d\pi P(0, T)^{\lfloor t/T \rfloor} P(0, t - \lfloor t/T \rfloor T)$$

This expression allows a computation of the transient distribution at any time $t \in \mathbb{R}^+$ without needing to integrate over ranges larger than the period T .

Example 4.2 Assume that $D_n(t) = (1 + \sin(\frac{2\pi}{T}t)) D_n$ for all $t \in \mathbb{R}_0^+$ and $n \in \mathbb{N}_0$, with $\Delta = (D_n : n \in \mathbb{N}_0)$ being the generating sequence of an homogeneous BMAP. Further, set the service rate to be constant in time, i.e. $\mu_t = \mu$ for all $t \in \mathbb{R}_0^+$. Denote the resulting generator at time 0 by $G := G(0)$. Then according to example A.4, the transient distribution is determined by

$$P(0, s) = \exp \left(\left(s + \frac{T}{2\pi} \left(1 - \cos \left(\frac{2\pi}{T} s \right) \right) \right) \cdot G \right)$$

for all $s \leq T$.

4.2.2. Stability and Asymptotic Distributions

In difference to $SMAP/M_t/c/c$ queues with finitely many phases and hence finite state space, the asymptotic analysis of the $BMAP/M_t/c$ queue first requires the derivation of an ergodicity condition, before asymptotic distribution can be given.

One of the most important criteria for ergodicity of Markov chains occurring in queueing theory is Foster's condition (see Foster [40], [41]). This has been extended in various directions afterwards (see Pakes [78], Mauldon [62], Marlin [61], Rosberg [84], Tweedie [90], [91] or Asmussen [4], p.18f). In Fayolle, Malyshev, Menshikov [39], a version of Foster's criterion for state spaces $\mathbb{N}_0 \times \{1, \dots, m\}$, hence for Markovian BMAP queues, can be found.

This shall be used in this section in order to prove a necessary and sufficient condition of ergodicity for the periodic $BMAP/M_t/c$ queue. After that, the periodic families of asymptotic distributions are given for the case of ergodicity.

Let $Q := (Q_t : t \in \mathbb{R}_0^+)$ denote a periodic $BMAP/M_t/c$ queue with period length $T > 0$. Denote the time-dependent generating sequence of the periodic BMAP by $(D_n(t) : n \in \mathbb{N}_0, t \in [0, T])$ and let μ_t denote the service rate at time t . Define the homogeneous Markov chain $Y = (Y_n : n \in \mathbb{N}_0)$ by $Y_n := Q_{nT}$ for all $n \in \mathbb{N}_0$. Furthermore, denote the phase process of Q , which coincides with that of the BMAP (N, J) , by J . Since the phase space is finite, the homogeneous Markov chain $(J_{nT} : n \in \mathbb{N})$ has an asymptotic and stationary distribution which shall be denoted by $\pi = (\pi_1, \dots, \pi_m)$. Finally, let $\mathbf{1}_m$ denote the m -dimensional column vector with all entries being 1 and denote the i th canonical row base vector of \mathbb{R}^m by e_i .

Theorem 4.8 Assume that the phase process J is irreducible. Then the Markov chain Y is ergodic if and only if the stability condition

$$\int_0^T \pi \sum_{n=1}^{\infty} n D_n(t) 1_m dt < c \cdot \int_0^T \mu_t dt \quad (4.4)$$

holds.

Proof: Let $A = (A_t : t \in \mathbb{R}_0^+)$ and $B = (B_t : t \in \mathbb{R}_0^+)$ denote the BMAP arrival process into Q and the Poisson process with rates $(\mu_t : t \in \mathbb{R}_0^+)$, respectively. That means, A_t is the random variable of all arrivals into the queue. Define $Z := A - B$ as the difference of these independent processes. Then Z is a periodic Markov jump process and by theorem 3.11, the mean expectation over one period length in phase equilibrium equals

$$E(Z_T) = E(A_T) - E(B_T) = \int_0^T \pi \sum_{n=1}^{\infty} n D_n(t) 1_m dt - c \cdot \int_0^T \mu_t dt \quad (4.5)$$

In order to show the necessity of condition 4.4, assume that $E(Z_T) \geq 0$. Since the state space of Z is $\mathbb{Z} \times \{1, \dots, m\}$, but the chain Y has a barrier at the zero level, we have

$$E(Y_1) > E(Z_T) \geq 0$$

for initial distributions with support $\{0\} \times \{1, \dots, m\}$. Starting in phase equilibrium, the asymptotic expectation

$$\lim_{n \rightarrow \infty} E(Y_n) = \lim_{n \rightarrow \infty} \sum_{k=1}^n E(Y_1) = \infty$$

diverges to infinity. Hence, there is no asymptotic distribution for Y .

Now we show sufficiency. Denote the transition probability matrix of the homogeneous Markov chain $(Z_{nT} : n \in \mathbb{N}_0)$ by p^Z . Since Z is homogeneous in the second component, we can define

$$p_k^Z(i, j) := p_{(0,i),(k,j)}^Z = P(Z_T = (k, j) | Z_0 = (0, i))$$

for all $k \in \mathbb{Z}$ and $i, j \in \{1, \dots, m\}$. The above observation 4.5 yields

$$\sum_{i=1}^m \pi_i \sum_{k \in \mathbb{Z}} \sum_{j=1}^m k \cdot p_k^Z(i, j) < 0$$

According to Fayolle, Malyshev, Menshikov [39], p.35, there is an $\varepsilon > 0$ and a positive function f such that

$$\sum_{n=0}^{\infty} \sum_{j=1}^m p_{(k,i),(n,j)}^Z \cdot f(n, j) - f(k, i) < -\varepsilon$$

4. Markovian Spatial Queues

for almost all states $(k, i) \in \mathbb{N}_0 \times \{1, \dots, m\}$. Furthermore, there are numbers a_1, \dots, a_m such that $f(k, i) = k + a_i$ for almost all states (k, i) .

Define the event

$$R(n) := \{\exists t \in [nT, (n+1)T[: Q_t = 0\}$$

for all $n \in \mathbb{N}_0$. Then the transition probabilities of the homogeneous chain Y can be decomposed in

$$\begin{aligned} p_{(k,i),(l,j)}^Y &:= P(Y_{n+1} = (l, j) | Y_n = (k, i)) \\ &= P(Y_{n+1} = (l, j) | Y_n = (k, i), R(n)) \cdot P(R(n) | Y_n = (k, i)) \\ &\quad + P(Y_{n+1} = (l, j) | Y_n = (k, i), R(n)^c) \cdot P(R(n)^c | Y_n = (k, i)) \\ &= P(Y_{n+1} = (l, j) | Y_n = (k, i), R(n)) \cdot P(R(n) | Y_n = (k, i)) \\ &\quad + p_{(k,i),(l,j)}^Z \cdot P(R(n)^c | Y_n = (k, i)) \end{aligned}$$

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

$$\lim_{k \rightarrow \infty} P(R(n) | Y_n = (k, i)) = 0$$

there is a $k_0 \in \mathbb{N}$ such that

$$P(R(n) | Y_n = (k, i)) < \frac{\varepsilon}{2 \cdot M'}$$

for all $k > k_0$, with $M' := M + \sum_{i=1}^m |a_i|$. Furthermore, condition 4.4 implies

$$\max_{i \in \{1, \dots, m\}} \int_0^T e_i \sum_{n=1}^{\infty} n D_n(t) 1_m dt < M < \infty$$

and thus the estimation

$$\begin{aligned} &\sum_{l=0}^{\infty} \sum_{j=1}^m P(Y_{n+1} = (l, j) | Y_n = (k, i), R(n)) \cdot f(n, j) \\ &\leq \sum_{l=0}^{\infty} \sum_{j=1}^m P(Y_{n+1} = (l, j) | Y_n = (k, i), R(n)) \cdot \left(l + \sum_{i=1}^m |a_i| \right) \\ &< M + \sum_{i=1}^m |a_i| \end{aligned}$$

holds. Using the positive function f , we now have

$$\begin{aligned} &\sum_{n=0}^{\infty} \sum_{j=1}^m p_{(k,i),(n,j)}^Y \cdot f(n, j) - f(k, i) \\ &< \sum_{n=0}^{\infty} \sum_{j=1}^m p_{(k,i),(n,j)}^Z \cdot f(n, j) - f(k, i) + \frac{\varepsilon}{2 \cdot M'} \cdot \left(M + \sum_{i=1}^m |a_i| \right) \\ &< -\frac{\varepsilon}{2} \end{aligned}$$

for almost all states $(k, i) \in \mathbb{N}_0 \times \{1, \dots, m\}$. Now Foster's criterion as stated in Fayolle, Malyshev, Menshikov [39], p.29, assures that Y is ergodic.
 ☺

In the case of Y being positive recurrent we can proceed as in the previous section in order to determine the asymptotic distribution. Thus we conclude this chapter with

Theorem 4.9 *If the stability condition 4.4 holds, then Q has a periodic family of asymptotic distributions assuming the form*

$$q_s = qP(0, s)$$

for all $s \in [0, T[$, with q being the asymptotic distribution of the embedded homogeneous Markov chain $Y = (Y_n : n \in \mathbb{N}_0)$ defined by $Y_n := Q_{nT}$ for all $n \in \mathbb{N}_0$.

Proof: This follows immediately from the above theorem 4.8 and from theorem A.28.

Example 4.3 Resuming example 4.2, we obtain $P(0, T) = e^{T \cdot G}$ and hence

$$qP(0, T) = q \iff qG = 0$$

for the distribution q . This and the formula for $P(0, s)$ in example 4.2 imply further that

$$q_s = q \exp \left(\left(s + \frac{T}{2\pi} \left(1 - \cos \left(\frac{2\pi}{T} s \right) \right) \right) \cdot G \right) = q$$

for all $s \in [0, t[$. Hence the queue given in this example has the same asymptotic behaviour as the homogeneous queue it is derived from.

4. *Markovian Spatial Queues*

5. Spatial Queues with Infinitely Many Servers

Spatial Queues with infinitely many servers arise naturally as models for the planning process of mobile communication networks. A very useful concept has been developed by Çinlar [30], partly on the basis of Massey and Whitt [63]. This assumes single arrivals distributed in time as a non-homogeneous Poisson process. Every arrival chooses a position in space according to a (time-dependent) distribution on the space and independent of all other users. Since the arrival process is a multi-dimensional Poisson process on the product space of time and arrival space, the queue process is Poisson as well, and the mean measures can be computed in a straightforward manner. This yields a nice solution to the queue process.

Mobile communication networks with parallel data streams of different types (e.g. voice, data, video) and group arrivals which are not independent in space require a more general treatment. As the BMAP-concept (see Lucantoni [60]) shows for non-spatial queueing models, parallel data types can be modeled by using phases. This will be done here, too. Furthermore, phases can be used to model special system characteristics. Spatial dependence in the positioning of group arrivals can be acknowledged by using Spatial Markovian Arrival Processes (SMAPs) as introduced in chapter 3.

In this chapter, spatial infinite server queues shall be examined in three stages of generality. Essentially, the same method of determining the queue process applies to all of them. In a first step (section 5.1), this method will be shown for the simplest model of homogeneous arrival rates in time and non-moving users. A first generalization in section 5.2 treats general (i.e. possibly non-homogeneous) arrival rates, but still non-moving users. In section 5.3, a last step provides for user movement as well general arrival rates.

5.1. Homogeneous Arrival Rates without User Movements

In this section, spatial queues with homogeneous arrival rates and constant user positions are examined. Let Q denote a queue with an SMAP arrival process (N, J) and infinitely many independent servers. Denote the arrival space by $(R, \sigma(R))$ and let $\{\Delta(S_1, \dots, S_n) : n \in$

5. Spatial Queues with Infinitely Many Servers

$\{IN, S_i \in \sigma(R)\}$ be the generating sequences determining (N, J) . Every incoming user is served immediately, i.e. there is no waiting time and the queue length is always zero.

For an introduction of the method of analysis, the service times will be assumed as general and iid in section 5.1.1, i.e. they do not depend on the user's position. In section 5.1.2, the model provides for service time distributions depending on the user's position.

5.1.1. IID Service Time Distributions

For ease of notation, the queue will be observed in only one subset $S \in \sigma(R)$ first. Observations in finitely many subsets will be derived in section 5.1.3 by the same method used in this section. Let $t \in \mathbb{R}^+$ denote the time instant and $S \in \sigma(R)$ the subset the queue is observed at. Further, let G denote the service time distribution function.

Define $Q_i(s, t; S)(y, A)$ as the probability of observing i users being served in S at time t and the system being in some phase $z \in A$ under the condition that at time $s < t$ there were no users being served in S and the system was in phase $y \in \Phi$. Let $Q_i(s, t; S)$ denote the kernel on Φ with entries $Q_i(s, t; S)(y, A)$. Further define the sequence $Q(s, t; S) := (Q_i(s, t; S) : i \in \mathbb{N}_0)$ and set $Q(t; S) := Q(0, t; S)$.

Remark 5.1 If $N(du; S)$ denotes the infinitesimal arrival rate in S at time u , then a stochastic integral describing the queue process at time t is given by

$$Q(t; S) = \int_0^t B(G^c(t-u), N(du; S))$$

with $G^c := 1 - G$ denoting the complement of the service time distribution function and $B((G^c(t-u), N(du; S)))$ being a random variable which is distributed binomially with success probability $G^c(t-u)$ and $N(du; S)$ trials. This integral is defined pathwise (cf. Liptser, Shiriyayev [59], p.90), i.e. for every $\omega \in \Omega$

$$\int_0^t B(G^c(t-u), N(du; S))(\omega) := \int_0^t B(G^c(t-u), N(du; S)(\omega))(\omega).$$

Although this integral form gives a short and nice representation of the queue, its evaluation is not easy. Therefore, another approach is taken for the computation of the transient distribution.

Fix any time $t \in \mathbb{R}_0^+$ the queue shall be observed at. In any infinitesimal time interval $du = \lim_{h \rightarrow 0}]u, u+h]$ with $u \in [0, t[$, batch arrivals occur with rates $\Delta(S) = (D_i(S) : i \in \mathbb{N}_0)$ according to the SMAP arrival process. Given that an arrival of batch size $m \in \mathbb{N}$ occurred during $]u, u+h]$ with $u \in [0, t[$, the probability of $i \in \{0, \dots, m\}$ arrivals still being served at time t is distributed binomially by

$$\binom{m}{i} G^c((t-u)-)^i G((t-u)-)^{m-i},$$

5.1. Homogeneous Arrival Rates without User Movements

denoting $G((t-u)-) := \lim_{h \rightarrow 0} G(t - (u+h))$ and $G^c(t) := 1 - G(t)$.

Conditioning upon the batch arrival rates, the total rate of i arrivals during a time interval du which are still in service at time t is

$$R_i(u, t; S) = \sum_{m=i}^{\infty} D_m(S) \binom{m}{i} G^c((t-u)-)^i G((t-u)-)^{m-i} \quad (5.1)$$

for every $u \in [0, t[$ and $i \in \mathbb{N}_0$.

The sequence $R(u, t; S) := (R_i(u, t; S) : i \in \mathbb{N}_0)$ can be interpreted as the time-dependent generating sequence of an inhomogeneous SMAP $H_{(t;S)} = (H_{(t;S)}(u) : u \leq t)$ which at time $u = t$ coincides with the infinite server queue that is to be examined. Note that for every time $t \in \mathbb{R}_0^+$ the queue is observed at, the sequence $R(u, t; S)$ and hence the SMAP $H_{(t;S)}$ is different. The transient distribution of the marginal queue process in the subset $S \in \sigma(R)$ at time $t \in \mathbb{R}_0^+$ is determined by $Q(t; S) = H_{(t;S)}(t)$. Formula A.2 yields the following representation:

Theorem 5.1 *The transient distribution of the marginal queue process of Q in the subset $S \in \sigma(R)$ at time $t \in \mathbb{R}_0^+$ is determined by*

$$Q(t; S) = \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_k} \dots \int_0^{u_2}}_{k \text{ integrals}} R(u_1, t; S) * \dots * R(u_k, t; S) du_1 \dots du_k \quad (5.2)$$

defining the case of zero integrals as the sequence $Id = (I, 0, 0, \dots)$ with I denoting the identity kernel and 0 the zero kernel on Φ , and further defining the convolution sequence $C = A * B$ by

$$C_n := \sum_{i=0}^n A_i B_{n-i}$$

for any two sequences A, B of kernels on Φ .

Proof: This was derived above.

A great disadvantage in the above formula 5.2 lies in the fact that the sequences $R(u, t; S)$ need to be determined separately for every time $t \in \mathbb{R}_0^+$ the queue shall be observed at. In the next theorem, a closed form for all times of observance is obtained:

Theorem 5.2 *The transient distribution of the marginal queue process of Q in the subset $S \in \sigma(R)$ at time $t \in \mathbb{R}_0^+$ can be written as*

$$Q(t; S) = \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_1} \dots \int_0^{u_{k-1}}}_{k \text{ integrals}} \tilde{R}(u_1; S) * \dots * \tilde{R}(u_k; S) du_k \dots du_1 \quad (5.3)$$

5. Spatial Queues with Infinitely Many Servers

with

$$\tilde{R}(u; S) := R(t - u, t; S)$$

being independent of $t > u$.

Proof: By equation 5.1, the rates

$$\tilde{R}_i(u; S) := R_i(t - u, t; S) = \sum_{m=i}^{\infty} D_m(S) \binom{m}{i} G^c(u-)^i G(u-)^{m-i}$$

are independent of $t > u$. Starting from equation 5.2, we have

$$\begin{aligned} Q(t; S) &= \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_k} \dots \int_0^{u_2}}_{k \text{ integrals}} R(u_1, t; S) * \dots * R(u_k, t; S) du_1 \dots du_k \\ &= \sum_{k=0}^{\infty} \int_0^t \int_{u_1}^t \dots \int_{u_{k-1}}^t R(u_1, t; S) * \dots * R(u_k, t; S) du_k \dots du_1 \\ &= \sum_{k=0}^{\infty} \int_0^t \int_0^{u_1} \dots \int_0^{u_{k-1}} R(t - u_1, t; S) * \dots * R(t - u_k, t; S) du_k \dots du_1 \end{aligned}$$

which proves the statement.

☺

Remark 5.2 The representation 5.3 is from the point of view that one looks backward in time from time t until time 0. Thus, equation 5.3 gives the transition kernel of a process running backward in time with the phase process still running forward.

Example 5.1 Assume that the service time distribution is deterministic, i.e. $G = \delta_s$ is the Dirac measure with support $s > 0$. Then

$$\tilde{R}_i(u; S) = \begin{cases} D_i(S) & \text{for } u < s \\ \delta_{i,0} \cdot D & \text{for } u \geq s \end{cases}$$

for all $i \in \mathbb{N}_0$, denoting the Kronecker function by δ and the generator of the phase process by $D = \sum_{n=0}^{\infty} D_n(S)$. Thus we obtain $\tilde{R}(u; S) = \Delta(S)$ for $u < s$ and $\tilde{R}(u; S) = D \cdot Id = (D, 0, 0, \dots)$ for $u \geq s$ and hence

$$Q(t; S) = \begin{cases} e^{\Delta(S) \cdot t} & \text{for } t < s \\ P^{\Phi}(0, t - s) e^{\Delta(S) \cdot s} & \text{for } t \geq s \end{cases}$$

for all $t \in \mathbb{R}_0^+$, with P^{Φ} denoting the transition probability kernel of the phase process.

The next two theorems yield expressions for the expectation kernel of the marginal queue process N at any time of observance.

5.1. Homogeneous Arrival Rates without User Movements

Theorem 5.3 Assume that the arrival rate in S is finite for any phase $y \in \Phi$, i.e.

$$\sum_{n=1}^{\infty} nD_n(S)(y, \Phi) < M < \infty \quad (5.4)$$

for all $y \in \Phi$. Then the expectation kernel of the marginal queue process of $Q = (N, J)$ in the subset $S \in \sigma(R)$ at time $t \in \mathbb{R}_0^+$ is given by

$$E(N_t) = \int_0^t P^\Phi(0, u) \sum_{n=1}^{\infty} nD_n(S)P^\Phi(u, t) G^c((t-u)-) du \quad (5.5)$$

with P^Φ denoting the transition kernel of the phase process.

Proof: Fix a time $t \in \mathbb{R}_0^+$. Since $Q(t; S)$ equals the distribution of the SMAP $H_{(t,S)}$ at time t , the expectation kernel at time t is given by theorem 3.10 as

$$\begin{aligned} E(N_t) &= \int_0^t P^\Phi(0, u) \sum_{n=1}^{\infty} nR_n(u, t; S)P^\Phi(u, t) du \\ &= \int_0^t P^\Phi(0, u) \sum_{n=1}^{\infty} n \sum_{m=n}^{\infty} D_m(S) \binom{m}{n} G^c((t-u)-)^n G^c((t-u)-)^{m-n} P^\Phi(u, t) du \end{aligned}$$

Abbreviating $p(u) := G^c((t-u)-)$ and $q(u) := G^c((t-u)-)$, we have

$$\begin{aligned} E(N_t) &= \int_0^t P^\Phi(0, u) \sum_{n=1}^{\infty} n \sum_{m=n}^{\infty} D_m(S) \binom{m}{n} p^n(u) q^{m-n}(u) P^\Phi(u, t) du \\ &= \int_0^t P^\Phi(0, u) \sum_{m=1}^{\infty} \sum_{n=1}^m nD_m(S) \frac{m!}{n! \cdot (m-n)!} p^n(u) q^{m-n}(u) P^\Phi(u, t) du \\ &= \int_0^t P^\Phi(0, u) \sum_{m=1}^{\infty} mD_m(S) p(u) \sum_{n=1}^m \frac{(m-1)!}{(n-1)! \cdot (m-n)!} p^{n-1}(u) q^{m-n}(u) P^\Phi(u, t) du \\ &= \int_0^t P^\Phi(0, u) \sum_{m=1}^{\infty} mD_m(S) G^c((t-u)-) P^\Phi(u, t) du \end{aligned}$$

since

$$\sum_{n=1}^m \frac{(m-1)!}{(n-1)! \cdot (m-n)!} p^{n-1}(u) q^{m-n}(u) = \sum_{n=0}^{m-1} \frac{(m-1)!}{n! \cdot (m-1-n)!} p^n(u) q^{m-1-n}(u)$$

sums up to 1.

☺

Theorem 5.4 If condition 5.4 holds and the phase process has stationary distribution π , then the expectation of the additive process of $Q(S)$ at time $t \in \mathbb{R}_0^+$ starting in phase equilibrium is

$$E_\pi(N_t, J_t \in \Phi) = \int_\Phi \pi(dy) \sum_{n=1}^{\infty} nD_n(S)(y, \Phi) \cdot \int_0^t G^c(u-) du \quad (5.6)$$

5. Spatial Queues with Infinitely Many Servers

Proof: This follows from the above theorem 5.3. Since $\pi P^\Phi(0, u) = \pi$ for all $u \in \mathbb{R}_0^+$ and $P^\Phi(u, t)(y, \Phi) = 1$ for all $y \in \Phi$ and $u < t \in \mathbb{R}_0^+$, we have

$$\begin{aligned} E_\pi(N_t, J_t \in \Phi) &= \int_0^t \int_\Phi \pi(dy) \sum_{n=1}^{\infty} n D_n(S)(y, \Phi) G^c((t-u)-) du \\ &= \int_\Phi \pi(dy) \sum_{n=1}^{\infty} n D_n(S)(y, \Phi) \cdot \int_0^t G^c(v-) dv \end{aligned}$$

after substituting $v := t - u$.
 \odot

5.1.2. Spatially Variable Service Time Distributions

Let $\{S_k : k \in \mathbb{N}\}$ with $S_k \in \sigma(R)$, $S_k \cap S_l = \emptyset$ for $k \neq l$, and $\bigcup_{k=1}^{\infty} S_k = R$ be a partition of the arrival space R . For every $k \in \mathbb{N}$, assign a service time distribution function G_k , which is valid on S_k .

Let $S \in \sigma(R)$ be measurable and $K \in \mathbb{N}$ such that $S \subset \bigcup_{k=1}^K S_k$. The queue process in S can be described as the convolution of the queue processes in $S \cap S_1, \dots, S \cap S_K$. Thus, for every $t \in \mathbb{R}^+$ the distribution $Q(t; S)$ of users in subset S at time t can be described in terms of the distributions $Q(t; S \cap S_k)$ of users in the subsets $S \cap S_k$ at time t . This results in

$$Q(t; S) = *_{k=1}^K Q(t; S \cap S_k)$$

Here, the generating sequences $\tilde{R}(t; S \cap S_k)$ which determine the sequences $Q(t; S \cap S_k)$ for $k \in \{1, \dots, K\}$ are determined as in the preceding section 5.1.1. The convolution $*$ is defined iteratively as $*_{k=1}^1 A_k := A_1$ and

$$*_{k=1}^{n+1} A_k := (*_{k=1}^n A_k) * A_{n+1}$$

for all $n \in \mathbb{N}$ and sequences A_1, \dots, A_{n+1} of kernels on Φ .

5.1.3. Common Distribution in Finitely Many Subsets

The same method as in section 5.1.1 can be applied in order to determine the joint (transient) distribution of the queue process in finitely many subsets of the arrival space R at a time $t \in \mathbb{R}^+$. Be $n \in \mathbb{N}$ and $S_k \in \sigma(R)$ for $k \in \{1, \dots, n\}$ measurable subsets. The joint distribution $Q(t; S_1, \dots, S_n)$ of users in service after time t shall be defined as the distribution of the \mathbb{N}_0^n -valued random variable indicating the number of users in $S_1 \times \dots \times S_n$ at time t .

Fix a time $t \in \mathbb{R}^+$ and subsets $S_1, \dots, S_n \in \sigma(R)$ the queue is to be observed at. In any infinitesimal time interval $du = \lim_{h \rightarrow 0}]u, u + h]$ with $u \in [0, t[$, batch arrivals occur with rate $\Delta(S_1, \dots, S_n) = (D_{i_1, \dots, i_n}(S_1, \dots, S_n) : i \in \mathbb{N}_0)$ according to the SMAP arrival process. Given that an arrival of batch size $(m_1, \dots, m_n) \in \mathbb{N}^n$ occurred during $]u, u + h]$, the

5.1. Homogeneous Arrival Rates without User Movements

probability of $(i_1, \dots, i_n) \in \prod_{k=1}^n \{0, \dots, m_k\}$ arrivals still being served in (S_1, \dots, S_n) at time t is distributed by

$$\prod_{k=1}^n \binom{m_k}{i_k} G^c((t-u)-)^{i_k} G((t-u)-)^{m_k-i_k},$$

again denoting $G((t-u)-) := \lim_{h \rightarrow 0} G(t - (u+h))$ and $G^c(t) := 1 - G(t)$.

Conditioning upon the batch arrival rates, the total rate of (i_1, \dots, i_n) arrivals into the family (S_1, \dots, S_n) of subsets during a time interval du which are still in service at time t is

$$\begin{aligned} & R_{i_1, \dots, i_n}(u, t; S_1, \dots, S_n) = \\ &= \sum_{m_1=i_1}^{\infty} \dots \sum_{m_n=i_n}^{\infty} D_{m_1, \dots, m_n}(S_1, \dots, S_n) \prod_{k=1}^n \binom{m_k}{i_k} G^c((t-u)-)^{i_k} G((t-u)-)^{m_k-i_k} \end{aligned}$$

for every $u \in [0, t[$ and $i_1, \dots, i_n \in \mathbb{N}_0$.

As in the one-dimensional case, the time-dependent (n-dimensional) sequence

$$R(u, t; S_1, \dots, S_n) := (R_{i_1, \dots, i_n}(u, t; S_1, \dots, S_n) : i_1, \dots, i_n \in \mathbb{N}_0)$$

can be interpreted as the generating sequence of an inhomogeneous SMAP $H_{(t; S_1, \dots, S_n)} = (H_{(t; S_1, \dots, S_n)}(u) : u \leq t)$ which at time $u = t$ coincides with the infinite server queue that is to be examined.

With definitions and arguments completely analogous to those in section 5.1.1, one derives the following result:

Theorem 5.5 *The transient distribution of the marginal queue process of Q in the subsets $S_1, \dots, S_n \in \sigma(R)$ at time $t \in \mathbb{R}_0^+$ is determined by*

$$Q(t; S_1, \dots, S_n) = \tag{5.7}$$

$$= \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_1} \dots \int_0^{u_{k-1}}}_{k \text{ integrals}} \tilde{R}(u_1; S_1, \dots, S_n) * \dots * \tilde{R}(u_k; S_1, \dots, S_n) du_k \dots du_1$$

with

$$\tilde{R}(u; S_1, \dots, S_n) := R(t-u, t; S_1, \dots, S_n)$$

being independent of $t > u$. The case of zero integrals shall be defined as the sequence $Id = (Id_{i_1, \dots, i_n} : i_1, \dots, i_n \in \mathbb{N}_0)$ with entries

$$Id_{i_1, \dots, i_n} := \begin{cases} I & \text{for } \sum_{k=1}^n i_k = 0 \\ 0 & \text{otherwise} \end{cases}$$

5. Spatial Queues with Infinitely Many Servers

and I (resp. 0) denoting the identity kernel (resp. the zero kernel) on Φ . Furthermore, the convolution sequence $C = A * B$ shall be defined by

$$C_{i_1, \dots, i_n} := \sum_{j_1=0}^{i_1} \cdots \sum_{j_n=0}^{i_n} A_{j_1, \dots, j_n} B_{i_1-j_1, \dots, i_n-j_n}$$

for any two sequences A, B of kernels on Φ .

5.1.4. Stability and Asymptotic Distribution

The asymptotic behaviour of the homogeneous $SMAP/G/\infty$ queue already is apparent in formulae 5.3 and 5.5 of the transient distributions and expectation kernels, respectively. The marginal expectation $E(N_t)$ remains finite for $t \rightarrow \infty$ if and only if the mean service time $E(G)$ as well as the mean arrival rate are finite. As will be shown in this section, these conditions are, as intuition would suggest, the main ingredients of a stability condition for the homogeneous $SMAP/G/\infty$ queue.

After first defining stability for homogeneous queues, the asymptotic distribution and marginal expectation of users in the queue are given for stable $SMAP/G/\infty$ queues. Finally, a necessary and sufficient stability condition is proven for the case of a countable phase space at the end of this section.

Definition 5.1 An homogeneous queue $Q = (Q_t : t \in \mathbb{R}^+)$ shall be called **stable** if the asymptotic distribution $q := \lim_{t \rightarrow \infty} Q_t$ does exist and the marginal asymptotic distribution of users in the queue has finite expectation.

Theorem 5.6 If $Q = (N, J)$ has an asymptotic distribution, it is determined by the kernel

$$Q(S) = \sum_{k=0}^{\infty} \underbrace{\int_0^{\infty} \int_0^{u_1} \cdots \int_0^{u_{k-1}}}_{k \text{ integrals}} \tilde{R}(u_1; S) * \cdots * \tilde{R}(u_k; S) du_k \cdots du_1$$

for every $S \in \sigma(R)$. Denote the asymptotic distribution of the phase process J by π . Then the asymptotic expectation of the marginal process N is given by

$$E(N(S), J \in \Phi) := \lim_{t \rightarrow \infty} E(N_t(S), J_t \in \Phi) = E(G) \cdot \int_{\Phi} \pi(dy) \sum_{n=1}^{\infty} n D_n(S)(y, \Phi)$$

for every $S \in \sigma(R)$, independent of the initial distribution.

Proof: The first statement follows immediately from formula 5.3 and the existence of the limit $Q(S) = \lim_{t \rightarrow \infty} Q(t; S)$. The second formula follows from equation 5.5 if one recognizes that with t growing to infinity, the term $G^c((t-u)-)$ tends to zero for the times u during which the phase process is not in equilibrium yet.

☺

Example 5.2 As in example 5.1, assume that the service times are deterministically s . Further assume that the phase process is asymptotic with asymptotic distribution π . Then the asymptotic distribution of Q under initial distribution ν is given by

$$Q(S) = \lim_{t \rightarrow \infty} \int d\nu Q(t; S) = \left(\lim_{t \rightarrow \infty} \int d\nu P^\Phi(0, t) \right) e^{\Delta(S) \cdot s} = \pi e^{\Delta(S) \cdot s}$$

for all $S \in \sigma(R)$.

Remark 5.3 In order to compute the asymptotic distribution of $Q = (N, J)$, it might be advantageous to use renewal theoretic methods as follows. Assume that Φ is countable, which is sufficient for most practical applications. Let $(T_n : n \in \mathbb{N}_0)$ denote the stopping times $T_0 := \inf\{t \in \mathbb{R}_0^+ : N_t = 0\}$ and $T_{n+1} := \inf\{t > T_n | N_t = 0, \exists s \in [T_n, t] : N_s \neq 0\}$ for all $n \in \mathbb{N}_0$. Further define $(X_n : n \in \mathbb{N}_0)$ by $X_n := J_{T_n}$ for all $n \in \mathbb{N}_0$. Then the discrete-time process (X, T) is a Markov-renewal process and Q is a semi-regenerative process (see Çinlar [26], pp.313,343). Hence one can use theorem 6.12 in Çinlar [26], p.347, to compute the asymptotic distribution of Q .

Assume without restriction of generality that $T_0 := 0$. If Q has an asymptotic distribution, then the process $X = (X_n : n \in \mathbb{N}_0)$ has an invariant probability measure ν . Define the mean values $m(j) := E(T_1 | X_0 = j)$ for all $j \in \Phi$ and let m denote the vector with entries $m(j)$. Further, denote

$$K_t(j, A) = P(Q_t \in A, T_1 > t | X_0 = j)$$

for all $t \in \mathbb{R}_0^+$, $j \in \Phi$ and $A \in \mathcal{S}$. The kernels K_t as well as the distribution ν and the vector m are determined as follows. Define $K_{t,u}^{(k)}(n)$ as the probability kernel on Φ that $T_1 > u$ and that $k \in \mathbb{N}_0$ batch arrivals in $S \in \sigma(R)$ are observed until time $u < t \in \mathbb{R}_0^+$, of which n single arrivals remain in the system until time t . Using the Kronecker function δ , we obtain

$$K_{t,u}^{(0)}(n) = \delta_{n,0} \cdot e^{D_0(S)u}$$

and iteratively

$$\begin{aligned} K_{t,u}^{(k+1)}(n) &= \int_0^u \sum_{m=0}^n K_{t,v}^{(k)}(n-m) \sum_{l \geq m, l \neq 0} D_l(S) \\ &\quad \cdot (G(t-v)^{l-m} G^c(t-v)^m - \delta_{n,0} \delta_{m,0} G(u-v)^l) e^{D_0(S)(u-v)} dv \end{aligned}$$

for all $u \leq t \in \mathbb{R}_0^+$, $n \in \mathbb{N}_0$ and $k \in \mathbb{N}_0$. Let $A = \{n\} \times B$ with $n \in \mathbb{N}_0$ and $B \in \sigma(\Phi)$. Then we have

$$K_t(j, A) = \sum_{k=0}^{\infty} K_{t,t}^{(k)}(n)(j, B)$$

5. Spatial Queues with Infinitely Many Servers

for all $j \in \Phi$. Define the kernel $P(T_1 > t)$ on Φ by

$$P(T_1 > t)(i, B) := \sum_{n=0}^{\infty} K_t(i, \{n\} \times B)$$

for all $i \in \Phi$ and $B \in \sigma(\Phi)$. Then the distribution ν is given as the solution of

$$\nu = \nu \int_0^{\infty} dP(T_1 = t)$$

Further, the mean vector m is given by its entries

$$m(i) = \int_0^{\infty} P(T_1 > t)(i, \Phi) dt$$

for all $i \in \Phi$.

Using the above values, the asymptotic distribution of Q is given by

$$q(A) = \lim_{t \rightarrow \infty} Q_t(A) = \frac{1}{\nu m} \sum_{j \in \Phi} \nu(j) \int_0^{\infty} K_t(j, A) dt$$

for all $A \in \mathcal{S}$, provided that the right-hand expression does exist.

In the rest of this section, a stability condition is derived for the case of the phase space Φ being finite.

Theorem 5.7 *Let $Q = (N, J)$ denote an homogeneous SMAP/G/ ∞ queue with finite phase space Φ . Assume that the phase process J is irreducible and has stationary distribution π . Denote the number of phases by m and the m -dimensional vector with all entries being 1 by 1_m . Then Q is stable if and only if the stability condition*

$$\pi \sum_{n=1}^{\infty} n D_n(R) 1_m \cdot E(G) < \infty \tag{5.8}$$

holds.

Proof: First we show necessity. Assume that the queue is stable. Then the asymptotic expectation of the marginal process $N(R)$ is given by

$$E(N(R), J \in \Phi) = E(G) \cdot \pi \sum_{n=1}^{\infty} n D_n(R) 1_m < \infty$$

according to the above theorem 5.6. This yields the necessity of condition 5.8.

5.1. Homogeneous Arrival Rates without User Movements

Sufficiency will be shown by comparison to the $M/G/\infty$ queue. Since Q is stable if and only if $Q(R)$ is stable, it suffices to refer to $Q(R)$ only. Denote the work load process of $Q(R)$ by $W = (W_t : t \in \mathbb{R}_0^+)$ and define

$$\gamma_{\max} := -\min_{i \in \Phi} D_0(R)(i, i) = \max_{i \in \Phi} |D_0(R)(i, i)|$$

as the maximal exit rate over all phases. Define \tilde{Q} as the following $M/G/\infty$ queue. The arrival process of \tilde{Q} shall be a Poisson process with rate γ_{\max} . Let $i_{\max} \in \Phi$ denote a phase with the highest expectation of the arrival batch size, i.e.

$$\frac{1}{\gamma_{i_{\max}}} \sum_{n=1}^{\infty} n D_n(R)(i_{\max}, \Phi) \geq \frac{1}{\gamma_j} \sum_{n=1}^{\infty} n D_n(R)(j, \Phi)$$

for all $j \in \Phi$. Since there are no absorbing states, the exit rate $\gamma_{i_{\max}}$ is a positive number. Define the service time distribution of \tilde{Q} by

$$\tilde{G} := \frac{1}{\gamma_{i_{\max}}} \sum_{n=1}^{\infty} D_n(R)(i_{\max}, \Phi) G^{*n}$$

denoting the service time distribution of Q by G and the n -fold convolution of G by G^{*n} . The $M/G/\infty$ queue constructed in this way has an arrival rate which is an upper bound of the phase specific arrival rates of $Q(R)$. Furthermore, the batch arrivals of $Q(R)$ are interpreted as single arrivals in \tilde{Q} with only one server working off the accrued service requirement of all the users in the batch arrival. Thus the work load $\tilde{W} = (\tilde{W}_t : t \in \mathbb{R}_0^+)$ of \tilde{Q} is certainly at least as high as the work load W of Q . Since by Wald's equation

$$E(\tilde{G}) = \frac{1}{\gamma_{i_{\max}}} \sum_{n=1}^{\infty} D_n(R)(i_{\max}, \Phi) \cdot n E(G) = \frac{1}{\gamma_{i_{\max}}} \cdot E(G) \cdot \sum_{n=1}^{\infty} n D_n(R)(i_{\max}, \Phi)$$

we have $E(\tilde{G}) < \infty$ by assumption of the stability condition. As $\gamma_{\max} < \infty$ is finite, too, \tilde{Q} is stable. This implies that \tilde{W} is positive recurrent. i.e. the expected duration between the time instants of \tilde{W} reaching the state 0 is finite. Since

$$\tilde{W}_t \geq W_t \geq 0$$

certainly for all $t \in \mathbb{R}_0^+$, the work load process W of Q is positive recurrent, too. Hence Q has an asymptotic distribution. The finiteness of the asymptotic expectation of the marginal process $N(R)$ is immediate from theorem 5.6 and the stability condition. \odot

Remark 5.4 The plausibility of condition 5.8 results from the obvious analogy to the stability condition for the resulting $BMAP/G/\infty$ queue on the set R . Of course, the $SMAP/G/\infty$ queue on $(R, \sigma(R))$ is stable if and only if the resulting $BMAP/G/\infty$ queue on the set R is stable.

5.2. General Arrival Rates without User Movements

In mobile communication networks, the typical characteristics of the arrival stream change in time periodically (e.g. over the course of the day). Stochastic models reflect this best by inhomogeneous arrival processes. The concept of SMAPs introduced in chapter 3 does provide for arrival rates varying in time and hence can be used to model this phenomenon. An examination analogous to the homogeneous case yields a solution for the corresponding spatial infinite server queue.

In this section, the same method of analysis as in the preceding section 5.1 shall be applied to spatial infinite server queues with general arrival rates. User movements will be examined in the next section. After the derivation of the transient distribution, an approximation method for the distribution at arbitrary time instants will be given.

5.2.1. Transient Distribution

The infinite server queue fed by an inhomogeneous SMAP can be analyzed analogously to the homogeneous case in section 5.1. Let

$$Q = (Q(t; S_1; \dots, S_n) : t \in \mathbb{R}_0^+, n \in \mathbb{N}, S_1; \dots, S_n \in \sigma(R))$$

denote a spatial queue with general SMAP input (N, J) , general service time distribution G and infinitely many servers. The arrival process (N, J) shall be defined by its time-dependent generating sequences $(\Delta(t; S_1; \dots, S_n) : t \in \mathbb{R}_0^+, n \in \mathbb{N}, S_1; \dots, S_n \in \sigma(R))$.

Again, the reason for restriction to a single subset and spatially constant service time distributions is ease of notation. The generalizations towards spatially variable service times and joint distributions in finitely many subsets are straightforward in the same way as in section 5.1. Be $t \in \mathbb{R}^+$ the time instant and $S \in \sigma(R)$ the subset the queue is observed at. While the integral representation of the queue process in remark 5.1 remains valid, the arrival rates at times $s \in [0, t]$ are not constant anymore, but depend on the time instant s .

The total rate of i arrivals during a time interval $du = \lim_{h \rightarrow 0}]u, u + h]$ is

$$R_i(u, t; S) = \sum_{m=i}^{\infty} D_m(u; S) \binom{m}{i} G^c((t-u)-)^i G((t-u)-)^{m-i}$$

for every $u \in [0, t[$ and $i \in \mathbb{N}_0$. As in the case of homogeneous arrival rates, the sequence $R(u, t; S) = (R_i(u, t; S) : i \in \mathbb{N}_0)$ can be interpreted as the time-dependent sequence of transition rates of an inhomogeneous SMAP $H_{(t;S)} = (H_{(t;S)} : u \leq t)$ which at time $u = t$ coincides with the infinite server queue that is to be examined.

The same arguments as in section 5.1.1 lead to the result

Theorem 5.8 *The transient distribution of the marginal queue process of Q in the subset $S \in \sigma(R)$ at time $t \in \mathbb{R}_0^+$ is determined by*

$$Q(t; S) = \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_k} \dots \int_0^{u_2}}_{k \text{ integrals}} R(u_1, t; S) * \dots * R(u_k, t; S) du_1 \dots du_k \quad (5.9)$$

defining the case of zero integrals as the sequence $Id = (I, 0, 0, \dots)$ with I denoting the identity kernel and 0 the zero kernel on Φ , and further defining the convolution sequence $C = A * B$ by

$$C_n := \sum_{i=0}^n A_i B_{n-i}$$

for any two sequences A, B of kernels on Φ .

Remark 5.5 Unfortunately, a unification of the above formula analogous to formula 5.3 cannot be achieved in general. This is due to the fact that

$$R(t-u, t; S) = \sum_{m=i}^{\infty} D_m(t-u; S) \binom{m}{i} G^c(u-)^i G(u-)^{m-i}$$

still depends on t if the arrival rates are inhomogeneous.

On the other hand, it is possible to extend the theory towards time-dependent service time distribution functions. This would lead to rates of the form

$$R_i(u, t; S) = \sum_{m=i}^{\infty} D_m(u; S) \binom{m}{i} G_u^c((t-u)-)^i G_u((t-u)-)^{m-i}$$

for every $u \in [0, t[$ and $i \in \mathbb{N}_0$, with G_u denoting the service time distribution function which holds for users arriving during the infinitesimal time interval du .

5.2.2. An Approximation

This section shows an approximation method which can be used in order to determine the distribution of the queue process at an arbitrary time $t \in \mathbb{R}_0^+$ without needing to integrate over the whole range $[0, t[$. The approximation regards only arrivals up to a certain time distance into the past. All arrivals which have occurred before are neglected. Thus, this method approximates the service time distribution by a distribution with cut tail.

Conditions for this approximation are the standard assumptions of finite mean service time and finite arrival rates. First, it is shown that the kernels of the generating sequence of $H_{(t;S)}$ are uniformly bounded. Then, this result is used to estimate the approximation error.

5. Spatial Queues with Infinitely Many Servers

Assume in this chapter that the mean service time

$$E(G) = \int_0^\infty G^c(u) du < \infty \quad (5.10)$$

is finite. Further assume that the arrival rate in the whole arrival space R

$$\sum_{m=1}^{\infty} m D_m(t; R)(y, \Phi) < M < \infty$$

is finite for all $t \in \mathbb{R}^+$ and $y \in \Phi$.

The next result gives the decisive bound for the approximation following.

Theorem 5.9 *For every $\varepsilon > 0$, there is a time distance $T(\varepsilon) \in \mathbb{R}^+$ such that the inequality*

$$\left\| \int_0^s R_n(u, t; S) du \right\| < M \cdot \varepsilon$$

holds for all $n \in \mathbb{N}_0$ and $s \leq t - T(\varepsilon)$, with $\|K\| := \sup_{y \in \Phi} K(y, \Phi)$ denoting the row norm for kernels on Φ .

Proof: Choose any $\varepsilon > 0$. By assumption 5.10, there is a $T(\varepsilon) \in \mathbb{R}^+$ such that

$$\int_{T(\varepsilon)}^\infty G^c(u-) du < \varepsilon$$

Fix any $y \in \Phi$ and $s \leq t - T(\varepsilon)$. Since all elements of $R_n(u, t; S)$ are positive for all $u \in [0, t]$ and $n \in \mathbb{N}$, we have

$$\left\| \int_0^s \sum_{n=1}^{\infty} R_n(u, t; S) du \right\| \leq \left\| \int_0^s \sum_{n=1}^{\infty} n R_n(u, t; S) du \right\|$$

Analogously to the proof of theorem 5.3, one can show that

$$\int_0^s \sum_{n=1}^{\infty} n R_n(u, t; S) du = \int_0^s \sum_{n=1}^{\infty} n D_n(u; S) G^c((t-u)-) du$$

Because $\sum_{n=1}^{\infty} n D_n(u; S) \leq \sum_{n=1}^{\infty} n D_n(u; R)$ for all $S \in \sigma(R)$ and $u \in \mathbb{R}_0^+$, the estimation

$$\begin{aligned} \left\| \int_0^s R_k(u, t; S) du \right\| &\leq \left\| \int_0^s \sum_{n=1}^{\infty} R_n(u, t; S) du \right\| \\ &\leq \left\| \int_0^s \sum_{n=1}^{\infty} n D_n(u; R) G^c((t-u)-) du \right\| \\ &\leq \int_0^s \left\| \sum_{n=1}^{\infty} n D_n(u; R) \right\| G^c((t-u)-) du \\ &< M \cdot \int_0^s G^c((t-u)-) du \end{aligned}$$

holds for all $k \in \mathbb{N}$. Substituting $v := t - u$ leads to

$$\begin{aligned} \left\| \int_0^s R_k(u, t; S) du \right\| &\leq \left\| \int_0^s \sum_{n=1}^{\infty} R_n(u, t; S) du \right\| < M \cdot \int_{t-s}^t G^c(v-) dv \\ &\leq M \cdot \int_{T(\varepsilon)}^{\infty} G^c(v-) dv < M \cdot \varepsilon \end{aligned}$$

Since

$$\int_0^s \sum_{n=1}^{\infty} R_n(u, t; S)(y, \Phi) du = - \int_0^s R_0(u, t; S)(y, \Phi) du$$

for all $s \in \mathbb{R}_0^+$, the proof is complete.

☺

Using this bound, we can prove the following approximation:

Theorem 5.10 *At any time $t \in \mathbb{R}^+$ and for every $y \in \Phi$, the distribution of the queue process can be approximated by*

$$Q(t; S)(y, \Phi) = Q(0, t; S)(y, \Phi) \approx Q(t - T(\varepsilon), t; S)(y, \Phi)$$

The approximation error is at most

$$|Q_n(t; S)(y, \Phi) - Q_n(t - T(\varepsilon), t; S)(y, \Phi)| < M \cdot \varepsilon$$

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

Proof: For every $k \in \mathbb{N}$, the difference between the respective k -fold integrals appearing in formula 5.9 is

$$\begin{aligned} &\int_0^t \dots \int_0^{u_{k-1}} R(u_k, t; S) * \dots * R(u_1, t; S) du_k \dots du_1 \\ &\quad - \int_{t-T(\varepsilon)}^t \dots \int_{t-T(\varepsilon)}^{u_{k-1}} R(u_k, t; S) * \dots * R(u_1, t; S) du_k \dots du_1 \\ &= \int_0^t \dots \int_0^{u_{k-2}} \int_0^{t-T(\varepsilon)} R(u_k, t; S) * \dots * R(u_1, t; S) du_k \dots du_1 \\ &= \int_0^{t-T(\varepsilon)} R(u_k, t; S) du_k * \int_0^t \dots \int_0^{u_{k-2}} R(u_{k-1}, t; S) * \dots * R(u_1, t; S) du_{k-1} \dots du_1 \end{aligned}$$

Summing up over all $k \in \mathbb{N}_0$ yields

$$\begin{aligned} Q(t; S) - Q(t - T(\varepsilon), t; S)(y, \Phi) &= \\ &= \int_0^{t-T(\varepsilon)} R(u, t; S) du * \sum_{k=0}^{\infty} \int_0^t \dots \int_0^{u_{k-1}} R(u_k, t; S) * \dots * R(u_1, t; S) du_k \dots du_1 \\ &= \int_0^{t-T(\varepsilon)} R(u, t; S) du * Q(t; S) \end{aligned}$$

5. Spatial Queues with Infinitely Many Servers

Hence, for any $y \in \Phi$ and $n \in \mathbb{N}_0$ we have

$$\begin{aligned} & |Q_n(t; S)(y, \Phi) - Q_n(t - T(\varepsilon), t; S)(y, \Phi)| \\ &= \left| \sum_{j=0}^n \int_{u=0}^{t-T(\varepsilon)} \int_{\Phi} R_j(u; S)(y, dz) du Q_{n-j}(t; S)(z, \Phi) \right| \\ &\leq \max_{j=0, \dots, n} \left| \int_0^{t-T(\varepsilon)} \int_{\Phi} R_j(u; S)(y, dz) du \right| \end{aligned}$$

since for every $z \in \Phi$ the probability $\sum_{j=0}^n |Q_{n-j}(t; S)(z, \Phi)|$ does not exceed 1. Because

$$\max_{j=0, \dots, n} \left| \int_0^{t-T(\varepsilon)} \int_{\Phi} R_j(u; S)(y, dz) du \right| = \max_{j=0, \dots, n} \left| \int_0^{t-T(\varepsilon)} R_j(u; S)(y, \Phi) du \right| < M \cdot \varepsilon$$

according to theorem 5.9, the proof is complete.

☺

5.3. General Arrival Rates with User Movements

User movements shall be modeled following the idea of Çinlar [30], pp.112f. Divide the space R into a finite partition $(S_1, \dots, S_K) \in (\sigma(R))^K$ with $K \in \mathbb{N}$, $\bigcup_{i=1}^K S_i = R$ and $S_i \cap S_j = \emptyset$ for $i \neq j$. Each set of this partition represents a spatial unit which is homogeneous in motion. Thus for times $s < t \in \mathbb{R}^+$, subsets $S_i, S \in \sigma(R)$ and $n, k \in \mathbb{N}_0$, one can define a probability $P(S_i, s, n; S, t, k)$ that under the condition that n users arrive in S_i at time s , k of them will be located in S at time t . At the time a user has finished using the network (i.e. its service time is over), it shall disappear and take on the position $\Upsilon \notin R$. This implies $P(S_i, s, n; R \cup \{\Upsilon\}, t, n) = 1$ for all $S_i \in \sigma(R)$, $s < t \in \mathbb{R}^+$ and $n \in \mathbb{N}_0$. Further, spatially variable service times can be subsumed by $G_i(t) = P(S_i, 0, 1; R, t, 0)$ or $\binom{m}{k} G_i^c(t)^k G_i(t)^{m-k} = P(S_i, 0, m; R, t, k)$ for batch arrivals. This convention allows a model of user movements which is analogous to the model for spatially variable service time distributions in section 5.1.2.

The transient distribution of the queue process can be calculated as a convolution over all $i \in \{1, \dots, K\}$ of the distributions of users which arrived in the set S_i until time t and assume a location in S at time t .

As in equation 5.1, the total rate of k users arriving in S_i during the infinitesimal time interval $du = \lim_{h \rightarrow 0}]u, u + h]$ and being located in S at time t is determined by

$$R_k(u, t; S_i, S) = \sum_{m=k}^{\infty} D_m(u; S_i) P(S_i, u+, m; S, t, k)$$

for every $u \in [0, t]$, $k \in \mathbb{N}_0$ and $S_i \in \sigma(R)$, denoting $u+ := \lim_{n \rightarrow \infty} u + \frac{1}{n}$.

5.3. General Arrival Rates with User Movements

Defining $R(u, t; S_i, S) := (R_k(u, t; S_i, S) : k \in \mathbb{N}_0)$ for every $u \in [0, t[$, the above rates determine the distributions $Q(t; S_i, S)$ of users arriving in S_i during du and being located in S at time t . The same arguments leading to formula 5.2 yield the expression

$$Q(t; S_i, S) = \sum_{k=0}^{\infty} \underbrace{\int_0^t \int_0^{u_k} \dots \int_0^{u_2} R(u_1, t; S_i, S) * \dots * R(u_k, t; S_i, S) du_1 \dots du_k}_{k \text{ integrals}}$$

The distribution of the queue as a convolution of these single distributions is

$$Q(t; S) = *_{i=1}^K Q(t; S_i, S)$$

An approximation of this expression is possible by the same idea as in section 5.2.2. In order to obtain a bound for the approximation error, we need the following

Theorem 5.11 *Assume that the mean service time is finite, i.e. that equation 5.10 holds. Be $\varepsilon > 0$ and determine $T(\varepsilon)$ as in theorem 5.10. Then the inequality*

$$\left\| \prod_{i=1}^K Q_{n_i}(0, t; S_i, S) - \prod_{i=1}^K Q_{n_i}(t - T(\varepsilon), t; S_i, S) \right\| < K \cdot M\varepsilon$$

holds for all $n_1, \dots, n_K \in \mathbb{N}_0$, with $\|K\| := \sup_{y \in \Phi} K(y, \Phi)$ denoting the row norm for kernels on Φ .

Proof: This is proven by induction. For $K = 1$ the statement is proven in the same way as in theorem 5.10. Assume the statement is true for any $K \geq 1$. Choose any $y \in \Phi$. To shorten the notation, define $p_i := Q_{n_i}(0, t; S_i, S)(y, \Phi)$ and $\tilde{p}_i := Q_{n_i}(t - T(\varepsilon), t; S_i, S)(y, \Phi)$. As probabilities, all p_i and \tilde{p}_i have values in $[0, 1]$. Then

$$\begin{aligned} \left| \prod_{i=1}^K p_i - \prod_{i=1}^K \tilde{p}_i \right| &= \left| \prod_{i=1}^{K-1} p_i (p_K - \tilde{p}_K) + \left(\prod_{i=1}^{K-1} p_i - \prod_{i=1}^{K-1} \tilde{p}_i \right) \tilde{p}_K \right| \\ &< \prod_{i=1}^{K-1} p_i \cdot M\varepsilon + (K-1) \cdot M\varepsilon \cdot \tilde{p}_K \end{aligned}$$

by induction hypothesis. Since $\prod_{i=1}^{K-1} p_i \leq 1$ and $\tilde{p}_K \leq 1$, the induction step is proven.

☺

Since the combinatorial possibilities of creating $n \in \mathbb{N}_0$ users out of K different sources grows with n , an error for the approximation of $Q_n(t; S)$ can only be given dependent on the number n of users.

Theorem 5.12 *The distribution of the queue process can be approximated by*

$$Q(0, t; S)(y, \Phi) \approx Q(t - T(\varepsilon), t; S)(y, \Phi)$$

5. Spatial Queues with Infinitely Many Servers

For every $i \in \mathbb{N}_0$, the approximation error is at most

$$|Q_n(0, t; S)(y, \Phi) - Q_n(t - T(\varepsilon), t; S)(y, \Phi)| < \binom{K+n-1}{K-1} \cdot K \cdot M\varepsilon$$

Proof: Using the same definitions for p_k and \tilde{p}_k as in the previous proof, the approximation error is

$$\begin{aligned} \left| \sum_{n_1+\dots+n_K=n} \prod_{k=1}^K p_k - \sum_{n_1+\dots+n_K=n} \prod_{k=1}^K \tilde{p}_k \right| &= \left| \sum_{n_1+\dots+n_K=n} \prod_{k=1}^K p_k - \prod_{k=1}^K \tilde{p}_k \right| \\ &< \sum_{n_1+\dots+n_K=n} K \cdot M\varepsilon \end{aligned}$$

by theorem 5.11. As can be proven by induction on K , the number of combinations of summing up $n_1 + \dots + n_K = n$ is $\binom{K+n-1}{K-1}$.

☺

5.4. Application: Planning a Mobile Communication Network

The results which have been obtained in this chapter can be applied to support the planning procedure of a mobile communication network. Assume that an area R is to be covered with base stations such as to provide network service for all users in it. This shall be achieved with a maximum outage probability of ε , which means that a user at any location in R , who wants to use the network, will find it busy with a probability of at most ε .

Then the analysis of spatial infinite server queues can help to derive a capacity function $C_\varepsilon : \sigma(R) \times \mathbb{R}^+ \rightarrow \mathbb{N}_0$ which yields for every subset $S \in \sigma(R)$ and every time $t \in \mathbb{R}^+$ a number $C_\varepsilon(S, t)$ such that if the network has the capacity of serving $C_\varepsilon(S, t)$ users in S at time t , then the quality of service with respect to the given maximum outage probability ε is guaranteed. Hence, in order to build a network that can guarantee the maximum outage probability ε , it suffices to provide service capacity according to the function C_ε . In practical applications, a finite partition (S_1, \dots, S_K) of R will be fine enough to answer all relevant questions concerning network planning. This implies that $\sigma(R)$ may be chosen as finite, too, which greatly simplifies computation of the function C_ε .

In order to determine C_ε , one can proceed as follows. Define an $SMAP/G/\infty$ queue with arrival space $(R, \sigma(R))$ on an appropriate level regarding time-dependence of arrival rates and user movements. An estimation procedure for the parameters of an SMAP can be found in the next chapter of this thesis. Then either the stationary distribution (in the homogeneous case) or the distributions at a given time of day (in the case of periodic arrival rates with period length being a day) can be represented by a set of functions $P_t : \sigma(R) \times \mathbb{N}_0 \rightarrow [0, 1]$ which

5.4. Application: Planning a Mobile Communication Network

yields for a given set $S \in \sigma(R)$ the probability $P_t(S, n)$ of n users being served in S at day time t . For the homogeneous case, we have a function P instead of a set $(P_t : t \in [0, 24])$ of functions which depend on the day time.

These are the distributions for an infinite server queue, which means that the probabilities were derived under the assumption that every user can be served. Hence the load for any real network with some positive outage probability would be lower. It thus suffices to determine the capacity function C_ε for the infinite server queue in order to find a sufficient capacity function for any real network. Since the outage probability ε usually will be chosen very small, the approximation by an infinite server queue seems reasonably sharp.

Now the capacity function for the infinite server queue can easily be determined as

$$C_\varepsilon(S, t) := \min \left\{ n \in \mathbb{N}_0 : \sum_{k=n+1}^{\infty} P_t(S, k) < \varepsilon \right\}$$

for all $t \in \mathbb{R}^+$ and $S \in \sigma(R)$. This means that the value $C_\varepsilon(S, t)$ of the capacity function at a subset $S \in \sigma(R)$ at a time t is given by the lowest number n of users such that the probability of more than n users being served in S would be smaller than ε for the respective infinite server queue. Since the load of any real network is lower than the load for the infinite server queue, the outage probability of a network satisfying this capacity function can be guaranteed to stay smaller than the given threshold ε .

5. *Spatial Queues with Infinitely Many Servers*

6. Parameter Estimation

In order to apply the concept of SMAPs and their respective queues, it is necessary to estimate the arrival parameters from real data streams. In general, this problem has not been solved yet. The problem of estimating parameters is more complex for SMAPs than for Batch Markovian Arrival Processes (BMAPs, see Lucantoni [60]), since homogeneous SMAPs generalize BMAPs towards a larger phase space and - most importantly - spatial batch arrivals. However, for certain subclasses of SMAPs estimation procedures can be derived.

Even for BMAPs a unified procedure for parameter estimation has not been developed yet. A survey of estimation methods is given in Asmussen [7]. His emphasis is on maximum likelihood estimation and its implementation via the EM algorithm. For the Markov Modulated Poisson Process (MMPP), an EM algorithm has been developed in Ryden [85], whereas Asmussen [6] contains a fitting procedure for phase-type distributions via the EM algorithm. An interesting different approach is given by Botta et al. [20] and Chauveau et al. [24]. They introduce the class of generalized hyperexponential distributions and derive statistical fitting methods by means of inversion of transforms.

In this chapter, an estimation procedure is introduced for the subclass of homogeneous SMAPs which have only finitely many phases and a finite σ -algebra $\sigma(R)$ on the arrival space R . For this class of SMAPs, the counting function of an arrival is reduced to a vector $C \in \mathbb{N}_0^K$ with $K = |\sigma(R)|$ denoting the number of elements in $\sigma(R)$. Since BMAPs belong to this subclass of SMAPs as a special case, the proposed estimation procedure yields a new result for BMAPs, too. This will be explicated in section 6.4.

Let m denote the number of phases of the SMAP which is to be estimated. The estimation procedure introduced in this chapter requires the measurement of the arrival time instants and the counting functions of the arrivals. Let $(T_n : n \in \{0, \dots, N\})$ denote the empirical arrival instants and $(C_n : n \in \{0, \dots, N\})$ the respective empirical counting functions of the spatial arrivals at times $(T_n : n \in \{0, \dots, N\})$. Define without loss of generality $T_0 := 0$.

The estimation works in three steps, each of which uses classical statistical methods. First, the empirical interarrival times are used to estimate the number m of phases and the matrix $D_0(R)$. Those can be interpreted as a sample of a phase-type distribution and hence $D_0(R)$ can be estimated by an EM-algorithm (see Dempster et al. [34] or for this special case Asmussen et al. [6]). If the number of phases is unknown, then the above maximum likelihood estimation can be repeated with increasing phase spaces until the likelihood gain does not exceed a certain threshold (cf. Jewell [52]). In a second step, for every empirical arrival instant the probability

6. Parameter Estimation

distribution of being in a certain phase at this instant is estimated using discriminant analysis (see Titterington et al. [89]). In the last step, the derived estimators of the first two steps are used in order to calculate the empirical estimator for the generator matrix. This is done according to standard estimators for Markov chains (see Anderson, Goodman [2]).

6.1. Arrival Rates

Let $(z_n := T_n - T_{n-1} : n \in \{1, \dots, N\})$ denote the empirical interarrival times. Assume first that the number of phases is known and denoted by m . According to theorem 3.8, the interarrival times of an SMAP are distributed phase-type with generator $D_0(R)$. Hence the $(z_n : n \in \{1, \dots, N\})$ are a sample of a phase-type distribution with density

$$z(t) = \pi e^{D_0(R)t} \eta(R)$$

for $t \in \mathbb{R}_+$. Here, $\pi = (\pi_1, \dots, \pi_m)$ is the steady-state distribution of the phase process at arrival instants and $\eta(R) := -D_0(R)1_m$ is the so-called exit vector of the phase-type distribution with representation $(\pi, D_0(R))$.

If m is known, there is a maximum likelihood estimator for π and $D_0(R)$. The solution of the estimating equations can be approximated iteratively by an EM algorithm (cf. Dempster et al. [34] or McLachlan, Krishnan [64]), which was derived for this special case by Asmussen et al. [6] and proceeds as follows.

For ease of notation, we write in this and the following section D_0 instead of $D_0(R)$ and η instead of $\eta(R)$. Denote the i th unit column vector by e_i and its transposition (i.e. the respective row vector) by e_i^T .

Starting from an intuitive first estimate $(\pi^{(1)}, D_0^{(1)})$ of the representation of the phase-type distribution, the recursions

$$\begin{aligned} \pi_i^{(k+1)} &= \frac{1}{N} \sum_{n=1}^N \frac{\pi_i^{(k)} b_i^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)} \\ D_{0;ij}^{(k+1)} &= \sum_{n=1}^N \frac{D_{0;ij}^{(k)} c_{ij}^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)} \bigg/ \sum_{i=1}^m \frac{c_{ii}^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)} \end{aligned}$$

for $i \neq j$ and

$$\eta_i^{(k+1)} = \sum_{n=1}^N \frac{\eta_i^{(k)} a_i^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)} \bigg/ \sum_{i=1}^m \frac{c_{ii}^{(k)}(z_n)}{\pi^{(k)} b^{(k)}(z_n)}$$

along with the relation

$$D_{0;ii}^{(k+1)} = -\eta_i^{(k+1)} - \sum_{j=1, j \neq i}^m D_{0;j}^{(k+1)}$$

and the definitions

$$\begin{aligned} a^{(k)}(z_n) &:= \pi^{(k)} e^{D_0^{(k)} z_n} \\ b^{(k)}(z_n) &:= e^{D_0^{(k)} z_n} \eta^{(k)} \\ c_{ij}^{(k)}(z_n) &:= \int_0^{z_n} \pi^{(k)} e^{D_0^{(k)} u} e_i e_j^T e^{D_0^{(k)}(z_n - u)} \eta \, du \end{aligned}$$

for $i, j \in \{1, \dots, m\}$ und $k \in \mathbb{N}$ lead to monotonically increasing likelihoods.

In Asmussen et al. [6], it is proposed to compute the values of $a^{(k)}(z_n)$, $b^{(k)}(z_n)$ and $c_{ij}^{(k)}(z_n)$ numerically as the solution to a linear system of homogeneous differential equations. According to their paper, this is possible achieving high precision by standard methods. The implementation by Gilbert [43] shows that, too.

The convergence of the EM algorithm is examined in Dempster et al. [34], chapter 3, as well as in McLachlan, Krishnan [64]. Improvements of the convergence rates are given in Meng, Dyk [66] and Jamshidian, Jennrich [51]. Titterington [88] gives an alternative recursion which is faster but not as stable.

A satisfying standard procedure for estimating the number m of phases has not been found yet. A feasible method without a prior estimation of m is proposed in Jewell [52]. Denote the maximum likelihood estimators of the representation of the phase-type distribution (as approximated by the EM algorithm described above) for an assumed number m_k of phases by $(\hat{\pi}^{(k)}, \hat{D}_0^{(k)})$. Further denote the resulting estimate of the exit vector by $\hat{\eta}^{(k)} := -\hat{D}_0^{(k)} \mathbf{1}_{m_k}$. Estimating the parameters by the above method for increasing m_k and stopping as soon as the likelihood gain

$$\prod_{n=1}^N \hat{\pi}^{(k+1)} e^{\hat{D}_0^{(k+1)} z_n} \hat{\eta}^{(k+1)} - \prod_{n=1}^N \hat{\pi}^{(k)} e^{\hat{D}_0^{(k)} z_n} \hat{\eta}^{(k)}$$

is smaller than a threshold value leads to a reasonable model fitting. Since the adaptation of the model increases with the assumed number of phases, the likelihood gain is always positive. The threshold value reflects the limit of accuracy beyond which the gain in model adaptation is not worth the additional computation time.

6.2. Phases at Arrival Instants

Using the estimator $(\hat{\pi}, \hat{D}_0)$ from the last section, the distribution of the non-observable phases at times $(T_n : n \in \mathbb{N})$ can be estimated using discriminant analysis in a standard way (cf. Titterington et al. [89], pp.168f).

6. Parameter Estimation

For a given empirical arrival instant T_n , let $Z_n := T_n - T_{n-1}$ and R_{n-1} denote the random variables of the last interarrival time and of the phase immediately after the last arrival instant T_{n-1} , respectively. Let $P(Z_n, R_{n-1})$ denote their joint distribution and $P(Z_n|R_{n-1})$, $P(R_{n-1}|Z_n)$ the conditional distributions.

Since $(Z_n : n \in \{1, \dots, N\})$ is given empirically by the time series $(T_n : n \in \{0, \dots, N\})$, it suffices to estimate the distribution $P(R_{n-1}|Z_n = z_n)$ for every $n \in \{1, \dots, N - 1\}$. This distribution is discrete, because by assumption there are only finitely many phases.

Bayes' formula yields for $j \in \{1, \dots, m\}$ and $n \in \{1, \dots, N\}$

$$P(R_{n-1} = j|Z_n = z_n) = \frac{P(Z_n = z_n|R_{n-1} = j) \cdot P(R_{n-1} = j)}{\sum_{i=1}^m P(Z_n = z_n|R_{n-1} = i) \cdot P(R_{n-1} = i)}$$

Since the interarrival times Z_n are distributed phase-type, the expressions $P(Z_n|R_{n-1})$ exist as conditional densities with respect to the Lebesgue measure on \mathbb{R} .

The expressions on the right hand can be estimated via the estimated parameters $(\hat{\pi}, \hat{D}_0)$ and the resulting vector $\hat{\eta}$. Hence, for every arrival instant T_{n-1} the estimator for the conditional distribution of the phase at T_{n-1} given the interarrival time $Z_n = T_n - T_{n-1}$ is given by

$$\hat{P}(R_{n-1} = i|Z_n = z_n) = \frac{e_i^T e^{\hat{D}_0 z_n \hat{\eta}} \cdot \hat{\pi}_i}{\sum_{j=1}^m e_j^T e^{\hat{D}_0 z_n \hat{\eta}} \cdot \hat{\pi}_j} = \frac{\hat{\pi}_i \cdot e_i^T e^{\hat{D}_0 z_n \hat{\eta}}}{\hat{\pi} e^{\hat{D}_0 z_n \hat{\eta}}}$$

for every $i \in \{1, \dots, m\}$ and $n \in \{1, \dots, N\}$.

Furthermore, it will be necessary to estimate the phase immediately before an arrival instant. This can be done by the same method. Denote the respective conditional distribution by $P(R_{n-}|Z_n)$. Then we get the estimation

$$\hat{P}(R_{n-} = i|Z_n = z_n) = \frac{\hat{\pi} e^{\hat{D}_0 z_n} e_i \cdot \hat{\eta}_i}{\sum_{j=1}^m \hat{\pi} e^{\hat{D}_0 z_n} e_j \cdot \hat{\eta}_j} = \frac{\hat{\pi} e^{\hat{D}_0 z_n} e_i \cdot \hat{\eta}_i}{\hat{\pi} e^{\hat{D}_0 z_n} \hat{\eta}}$$

for every $i \in \{1, \dots, m\}$ and $n \in \{1, \dots, N\}$.

6.3. Generator Matrix

Since the phase space $\Phi = \{1, \dots, m\}$ as well as the σ -algebra $\sigma(R)$ is finite, the infinitesimal transition rates are determined by the values for

$$D(i, j, C) = \lambda_i \cdot p_i(j, C)$$

for every $i, j \in \{1, \dots, m\}$ and $C \in IN_0^K$.

Denote the zero vector in IN_0^K by $\mathbf{0}$. In order to complete the estimation of the generator matrix, it suffices to estimate the parameters $p_i(j, C)$ for $C \neq \mathbf{0}$. Without any further assumptions regarding these, use of the empirical estimator is standard. This is given in Anderson,

Goodman [2]. Since in the present statistical model the phase process is hidden, the phase at each arrival instant cannot be observed but must be estimated. For this, the results of the last section are used.

For every $k \in \{1, \dots, N-1\}$, let R_k and R_{k-} denote the random variable of the (non-observable) phase at the empirical arrival instant T_k and immediately before it, respectively. In the last section, estimators for the conditional distributions of the R_k and R_{k-} were given. Let δ denote the Kronecker function for testing equality of counting functions and remember that C_k denotes the empirical counting function of the k th spatial arrival. Define

$$N_i(j, C) := \sum_{k=1}^{N-1} \hat{P}(R_{k-} = i | Z_k = z_k) \cdot \delta_{C_k, C} \cdot \hat{P}(R_k = j | Z_{k+1} = z_{k+1})$$

and

$$N_i := \sum_{k=1}^{N-1} \hat{P}(R_{k-} = i | Z_k = z_k) = \sum_{C \in \mathcal{N}_0^K} \sum_{j=1}^m N_i(j, C)$$

for $C \in \mathcal{N}_0^K$ and $i, j \in \{1, \dots, m\}$. In an SMAP, the random variables R_{k-} and R_k are dependent in any non-trivial case. Since in the present model the phases are non-observable, this dependency can only be reflected by conditioning on the consecutive empirical interarrival times z_k and z_{k+1} .

Because of

$$p_i(j, C) = \frac{P(R_{k-} = i, C_k = C, R_k = j)}{P(R_{k-} = i, C_k \neq \mathbf{0})} \cdot \frac{P(R_{k-} = i, C_k \neq \mathbf{0})}{P(R_{k-} = i)}$$

for all $k \in \{1, \dots, N-1\}$, the empirical estimator for $p_i(j, C)$ is given by

$$\hat{p}_i(j, C) := \frac{N_i(j, C)}{N_i} \cdot \frac{\hat{\eta}_i}{-\hat{D}_0(i, i)}$$

for every $C \in \mathcal{N}_0^K \setminus \{\mathbf{0}\}$ and $i, j \in \{1, \dots, m\}$.

For applications in the modelling of mobile communication networks, it usually suffices to set $\sigma(R) := \sigma(\{S_1, \dots, S_k\})$ with S_1, \dots, S_k being a finite partition of R , i.e. $R = \sum_{i=1}^k S_i$ and $S_i \cap S_j = \emptyset$ for $i \neq j$. As the discrete σ -algebra over a finite set of subsets, $\sigma(R)$ will then be finite, too. In such an application field, the subsets S_i describe entities with roughly homogeneous call behaviour like big buildings, housing blocks, roads, landscape areas etc.

6.4. Estimation for BMAPs

For the special case that the arrival space R consists of only one element, i.e. for the non-spatial version of the SMAPs considered in this chapter, we have BMAPs since the phase

6. Parameter Estimation

space is finite. Hence, the estimation procedure described above yields a method of parameter estimation for BMAPs as well.

While the first two steps of the procedure are exactly the same for the special case of BMAPs, the notation can be simplified in the third step. Instead of counting functions over the arrival space, the empirical data for BMAPs would only yield the size of batches which arrive in the whole (queueing) system. Thus, instead of counting functions $(C_k : k \in \{0, \dots, N\})$ it suffices to measure the sizes $(b_k : k \in \{0, \dots, N\})$, $b_i \in \mathbb{N}$, of batch arrivals into the system.

The generator matrix assumes the easier form

$$Q = \begin{pmatrix} D_0 & D_1 & D_2 & D_3 & \dots \\ & D_0 & D_1 & D_2 & \dots \\ & & D_0 & D_1 & \ddots \\ & & & D_0 & \ddots \\ & & & & \ddots \end{pmatrix}$$

with $(m \times m)$ -matrices

$$D_{n,ij} = \lambda_i \cdot p_i(n, j)$$

Defining

$$n_i(n, j) := \sum_{k=1}^{N-1} \hat{P}(R_{k-} = i | Z_k = z_k) \cdot \delta_{b_k, n} \cdot \hat{P}(R_k = j | Z_{k+1} = z_{k+1})$$

and

$$n_i := \sum_{k=1}^{N-1} \hat{P}(R_{k-} = i | Z_k = z_k) = \sum_{n=1}^{\infty} \sum_{j=1}^m n_i(n, j)$$

for $n \in \mathbb{N}$ and $i, j \in \{1, \dots, m\}$, the empirical estimator for the infinitesimal transition rates $(p_i(n, j) : n \in \mathbb{N}, i, j \in \{1, \dots, m\})$ can be written as

$$\hat{p}_i(n, j) = \frac{n_i(n, j)}{n_i} \cdot \frac{\hat{\eta}_i}{-\hat{D}_0(i, i)}$$

A test of the described estimation procedure can be obtained as follows. First, a time series is generated by a so-called input BMAP with given and known parameters. This time series serves as the data for the estimation procedure. Then the BMAP that is estimated via the procedure is compared to the input BMAP. The following results have been computed on a Pentium III computer with 500 MHz and 128 MB RAM. The length of the time series was always $N = 10,000$ arrivals.

In order to judge the quality of the estimations, two facts need to be considered. The Phase-type distribution as well as BMAPs are over-parametrized (cf. Botta et al. [20]), which implies

that two BMAPs with different parameters can have the same distribution. Hence for a given input BMAP it is not always possible to distinguish a single best estimation. Furthermore, the proposed estimation procedure is based on maximum likelihood estimation. Thus the estimated BMAP should be compared to the input BMAP by the ratio R of their respective likelihoods of the generated time series. This ratio with respect to the number N of arrivals in the time series (i.e. $R^{1/N}$) is given in each table as well as the computation time for the estimation.

The first two tables (see page 82) give typical examples of input BMAPs and the respective estimate. It can be seen that although the parameters differ, the likelihood of the generated time series in table 1 is even higher for the estimation than for the input BMAP.

The next two tables (see page 83) show the main observable weakness of the estimation procedure. Namely, if many parameters of the input BMAP are zero, then these are not estimated as zero and the likelihood of the time series for the estimated BMAP remains short of the likelihood for the input BMAP.

Finally, the last two tables (page 84) show that the estimated number of phases does not always coincide with the number of phases of the input BMAP.

As can be seen, the run times are very short. This does not change for BMAPs of more than two phases. Usually, a likelihood ratio per arrival of more than 0.90 can be achieved. For a detailed description of the implementation and many more numerical examples, see Gilbert [43].

6.5. Remarks

Remark 1: A straightforward method of deriving a maximum likelihood estimator for SMAPs (and thus for BMAPs, too) is an immediate application of the EM algorithm on the transition rates $(D_n(S)(i, j) : S \in \sigma(R), n \in \mathbb{N}_0, i, j \in \{1, \dots, m\})$ of the arrival process. This can be done using the maximum likelihood estimators for Markov processes derived in Albert [1].

However, the critical point of the estimation procedure are the run time and the storage requirements of the EM algorithm (see Meng, Dyk [66]). The straightforward method of estimation via an EM algorithm would apply this algorithm to a much larger set of parameters, which would lead to larger run time as well as storage requirements. The number of parameters which are to be determined iteratively via the EM algorithm is greatly reduced by dividing the estimation procedure into three steps as described in the sections above. It seems reasonable to expect considerable improvements in run time performance, if the statistical model that is to be estimated by the EM algorithm is as simple as possible.

Remark 2: The model which is examined in this paper is a special case of a hidden Markov model (see e.g. Leroux [58]). Here, the hidden Markov chain is the phase process at arrival instants and the observed variables are the consecutive interarrival times and the sizes of the batch arrivals. These are conditionally independent given the phase process. The conditional

6. Parameter Estimation

Table 1

	Input BMAP		Estimated BMAP	
$D_0 =$	-1.000	0.100	-1.046	0.171
	3.300	-10.000	3.376	-8.958
$D_1 =$	0.400	0.050	0.265	0.174
	1.650	1.700	1.499	1.067
$D_2 =$	0.200	0.025	0.136	0.088
	0.825	0.850	0.778	0.543
$D_3 =$	0.120	0.015	0.075	0.049
	0.495	0.510	0.461	0.330
$D_4 =$	0.072	0.009	0.047	0.031
	0.297	0.306	0.289	0.203
$D_5 =$	0.008	0.001	0.006	0.004
	0.033	0.034	0.031	0.021

run time: 34 seconds

likelihood ratio per arrival: 1.005

Table 2

	Input BMAP		Estimated BMAP	
$D_0 =$	-5.000	0.100	-5.048	0.770
	3.300	-65.000	11.691	-62.871
$D_1 =$	1.100	0.400	0.623	0.701
	11.200	7.400	6.721	8.453
$D_2 =$	0.200	2.000	0.387	1.333
	6.800	2.100	4.237	8.094
$D_3 =$	0.100	0.000	0.140	0.133
	12.100	4.300	4.907	5.072
$D_4 =$	0.600	0.200	0.315	0.346
	7.800	3.500	4.211	4.590
$D_5 =$	0.050	0.250	0.073	0.228
	1.300	5.200	1.229	3.665

run time: 19 seconds

likelihood ratio per arrival: 0.941

Table 3

	Input BMAP		Estimated BMAP	
$D_0 =$	-1.000	0.000	-1.053	0.168
	0.000	-10.000	3.554	-9.615
$D_1 =$	0.900	0.100	0.529	0.356
	3.300	6.700	3.283	2.777

run time: 36 seconds

likelihood ratio per arrival: 0.973

Table 4

	Input BMAP		Estimated BMAP	
$D_0 =$	-1.000	0.000	-1.032	0.154
	0.000	-10.000	1.722	-9.903
$D_1 =$	0.000	1.000	0.216	0.662
	10.000	0.000	3.755	4.426

run time: 10 seconds

likelihood ratio per arrival: 0.746

distributions of the interarrival times given the phase process belong to the family of phase-type distributions.

6. Parameter Estimation

Table 5

	Input BMAP				Estimated BMAP	
$D_0 =$	-1.000	0.100	0.010	0.090	-868.7	308.8
	1.000	-10.00	3.000	4.000	0.170	-1.130
	25.00	25.00	-100.0	25.00		
	499.0	1.000	1.000	-1000		
$D_1 =$	0.300	0.050	0.010	0.050	101.9	173.7
	0.250	0.250	0.250	0.250	0.150	0.330
	5.500	5.500	0.500	1.000		
	124.5	0.250	0.250	124.5		
$D_2 =$	0.150	0.030	0.000	0.020	51.34	85.58
	0.130	0.130	0.130	0.130	0.080	0.170
	2.750	2.750	0.250	0.500		
	62.25	0.130	0.130	62.25		
$D_3 =$	0.090	0.020	0.000	0.010	34.44	54.29
	0.080	0.080	0.080	0.080	0.040	0.010
	1.650	1.650	0.150	0.300		
	37.35	0.080	0.080	37.35		
$D_4 =$	0.050	0.010	0.000	0.010	18.99	33.28
	0.050	0.050	0.050	0.050	0.020	0.060
	0.990	0.990	0.090	0.180		
	22.41	0.050	0.050	22.41		
$D_5 =$	0.010	0.000	0.000	0.000	2.420	3.890
	0.010	0.010	0.010	0.010	0.000	0.010
	0.110	0.110	0.010	0.020		
	2.490	0.010	0.010	2.490		

run time: 29 seconds

likelihood ratio per arrival: 1.034

Table 6

	Input BMAP		Estimated BMAP		
$D_0 =$	-105.0	5.000	-107.3	4.075	16.49
	2.000	-52.00	3.160	-89.32	11.28
			4.310	3.051	-51.76
$D_1 =$	100.0	0.000	47.73	20.99	18.05
	0.000	50.00	41.12	18.10	15.66
			24.24	10.71	9.445

run time: 449 seconds

likelihood ratio per arrival: 0.988

Appendix

A. Markov Jump Processes

In queueing theory, Markov jump processes play a very important role. All Markovian queues can be modeled by Markov jump processes. Furthermore, since Doob's observation (cf. Doob [36], p.401-403) that the inter-arrival times of calls by many independent subscribers to a telephone network are distributed exponentially, Markov jump processes have become the most important type of arrival processes in queueing models.

As a main foundation for the rest of the present thesis and for convenience of reference, this chapter collects some classical results for (inhomogeneous) Markov jump processes with a general state space. Following mainly Gikhman, Skorokhod [44], the first section provides a sound definition, the most important sample path properties and a canonical way of constructing Markov jump processes. The second section treats the special case for which the generators satisfy a property called quasi-commutability. The chapter concludes with a short section exploring the case of periodic transition rates.

Markov Jump Processes will be used in this thesis in order to define Markov-additive jump processes in chapter 2 and, based on these, Spatial Markovian Arrival Processes in chapter 3. Further, an analysis of (periodic) Markovian queues will be performed using the results from this chapter.

A.1. Definition and Properties

In this section, Markov jump processes will be defined as Markov processes with transition probabilities that satisfy certain regularity conditions. These allow the definition of unique transition rates and lead to first elementary results. The most important results are the Kolmogorov differential equations, which allow the derivation of the transition probabilities. Since these determine the finite-dimensional marginal distribution of the process and hence the distribution of the process itself, the section can be concluded with the assertion that the definition at the beginning leads to a unique distribution of the process.

Definition A.1 Be $X = (X_t : t \in \mathbb{R}_0^+)$ a Markov process with a separable locally compact metric state space S and transition probabilities $P(s, t; x, A) := P(X(t) \in A | X(s) = x)$. Assume that the state space S has a σ -algebra \mathcal{S} with $\{x\} \in \mathcal{S}$ for every $x \in S$. The process X shall be called **Markov jump process** if the following conditions hold:

A. Markov Jump Processes

1. The limits

$$q(t; x, A) := \lim_{h \downarrow 0} \frac{P(t, t+h; x, A) - 1_A(x)}{h}$$

exist uniformly with respect to (t, x, A) .

2. The function $t \rightarrow q(t; x, A)$ is uniformly continuous on $S \times S$.

For every time $t \in \mathbb{R}_0^+$, state $x \in S$ and state set $A \in \mathcal{S}$, the limit $q(t; x, A)$ shall be called **infinitesimal transition rate** from x to A at time t . The kernel $Q(t)$ on S defined by $Q(t)(x, A) := q(t; x, A)$ shall be called **generator** of X at time t .

If X is time-homogeneous, i.e. if $q(t; x, A) = q(x, A)$ is constant in $t \in \mathbb{R}_0^+$, then the process is called **homogeneous Markov jump process**.

Remark A.1 Define a norm on the space of all kernels on S by

$$\|K\| := \sup_{x \in S, A \in \mathcal{S}} |K(x, A)|$$

for any kernel K . Denote the transition probability kernel from time s to time $t > s$ by $P(s, t)$. Then the conditions of the above definition can be written as follows:

1. The limits

$$Q(t) := \lim_{h \downarrow 0} \frac{P(t, t+h) - Id}{h}$$

exist uniformly with respect to t .

2. The function $t \rightarrow Q(t)$ is continuous.

Remark A.2 The first condition leads to the forward Kolmogorov differential equations (see theorem A.4). The second condition guarantees that these can be solved via the iteration method by Picard and Lindelöf (see Kamke [53], S.38). Then the transition probability kernels can be expressed in terms of the generators as will be shown in this section.

Remark A.3 Although the infinitesimal rates are defined via transition probabilities, in practice it is not necessary to know the latter in order to determine the former. In order to construct a Markov jump process, it suffices to define values for the infinitesimal rates which satisfy certain conditions. A way of constructing a Markov jump process is described at the end of this section.

The infinitesimal transition rate $q(t; x, A)$ can be regarded as the tendency of the process at time $t \in \mathbb{R}_0^+$ to jump from a state $x \in S$ to some state $y \in A \in \mathcal{S}$. The greater the infinitesimal transition rate, the higher the probability that the process performs this jump in an interval of time. For transition rates of ever greater size, this can lead to so-called explosions which means an accumulation point of jump times. The above definition excludes explosions and highly irregular behaviour, as is seen in

Theorem A.1 *The above definition implies the boundedness of the infinitesimal transition rates, i.e. there is an $M < \infty$ with*

$$|q(t; x, A)| < M$$

for all $t \in \mathbb{R}_0^+$, $x \in S$ and $A \in \mathcal{S}$.

Proof: The uniform convergence implies that there is a $\delta > 0$ such that

$$\left| \frac{P(t, t+h; x, A) - 1_A(x)}{h} - q(t; x, A) \right| < 1$$

for all $t \in \mathbb{R}_0^+$, $x \in S$, $A \in \mathcal{S}$ and $h < \delta$. Fix such an $h < \delta$. Since

$$|P(t, t+h; x, A) - 1_A(x)| \leq 1$$

the estimation

$$|q(t; x, A)| < 1 + \left| \frac{P(t, t+h; x, A) - 1_A(x)}{h} \right| \leq 1 + \frac{1}{h} =: M$$

holds.

☺

The definition of a Markov jump processes immediately yields some first properties of the infinitesimal rates and gives reason for some further definitions. In the following, the difference of two finite measures on the same measurable space shall be called a **charge**.

Theorem A.2 *For every $t \in \mathbb{R}_0^+$ and $x \in S$, the infinitesimal rate $q(t; x, A)$ as a function in A is a charge on \mathcal{S} . Further the properties*

$$q(t; x, S) = 0$$

$$q(t; x, A) = \lim_{h \downarrow 0} \frac{P(t, t+h; x, A) - 1_A(x)}{h} \geq 0$$

for $x \notin A$, and finally

$$q(t; x, \{x\}) = -q(t; x, S \setminus \{x\}) \leq 0$$

hold.

Proof: see Gikhman, Skorokhod [44], p.312

Definition A.2 For every $t \in \mathbb{R}_0^+$ and $x \in S$, define the values

$$\gamma(t, x) := -q(t; x, \{x\})$$

A. Markov Jump Processes

and

$$\gamma(t, x, A) := q(t; x, A \setminus \{x\})$$

as well as

$$p(t, x, A) := \begin{cases} \frac{\gamma(t, x, A)}{\gamma(t, x)} & \text{for } \gamma(t, x) > 0 \\ 1_A(x) & \text{for } \gamma(t, x) = 0 \end{cases}$$

for all $A \in \mathcal{S}$.

Theorem A.3 For fixed $(t, x) \in \mathbb{R}_0^+ \times S$, the above defined $p(t, x, A)$ as a function in A is a probability measure on \mathcal{S} with support set $S \setminus \{x\}$. For $\gamma(t, x) > 0$, the representation

$$p(t, x, A) = \lim_{h \downarrow 0} \frac{P(t, t+h; x, A \setminus \{x\})}{P(t, t+h; x, S \setminus \{x\})}$$

holds for all $A \in \mathcal{S}$.

Proof: For $\gamma(t, x) = 0$, $p(t, x, A) = 1_A(x)$ is a probability measure. Let $\gamma(t, x) > 0$. By theorem A.2, $q(t; x, A)$ is a charge on \mathcal{S} and $q(t; x, A) \geq 0$ for $x \notin A$. This means that $\gamma(t, x, A)$ is a finite measure, since $q(t; x, A)$ is bounded according to theorem A.1. Hence, $p(t, x, A) = \frac{\gamma(t, x, A)}{\gamma(t, x)}$ is a probability measure, because $q(t; x, \{x\}) = -q(t; x, S \setminus \{x\})$ by theorem A.2. The representation as a limit results from

$$\begin{aligned} p(t, x, A) &= \frac{\gamma(t, x, A)}{\gamma(t, x)} = \frac{q(t; x, A \setminus \{x\})}{-q(t; x, \{x\})} = \frac{q(t; x, A \setminus \{x\})}{q(t; x, S \setminus \{x\})} \\ &= \left(\lim_{h \downarrow 0} \frac{P(t, t+h; x, A \setminus \{x\})}{h} \right) \cdot \left(\lim_{h \downarrow 0} \frac{P(t, t+h; x, S \setminus \{x\})}{h} \right)^{-1} \\ &= \lim_{h \downarrow 0} \frac{P(t, t+h; x, A \setminus \{x\})}{P(t, t+h; x, S \setminus \{x\})} \end{aligned}$$

☺

Remark A.4 Define $P(t)$ by $P(t)(x, A) := p(t; x, A)$ for all $t \in \mathbb{R}_0^+$, $x \in S$ and $A \in \mathcal{S}$. Since $q(t; x, A)$ is uniformly continuous in (x, A) by definition and $p(t; x, \cdot)$ is a probability measure on \mathcal{S} by theorem A.3, $P(t)$ is a probability kernel on S for all $t \in \mathbb{R}_0^+$. $P(t)$ shall be called **jump kernel** of X at time t .

The most important results of this chapter are the so-called Kolmogorov differential equations. They state that the transition probabilities are differentiable backward and forward in the following sense:

Theorem A.4 For every $s \in \mathbb{R}_0^+$, $x \in S$ and $A \in \mathcal{S}$, the transition probability $P(s, t; x, A)$ is differentiable with respect to t , and

$$\begin{aligned} \frac{\partial P(s, t; x, A)}{\partial t} &= \int_S q(t, y, A) P(s, t; x, dy) \\ &= - \int_S \gamma(t, y) P(s, t; x, dy) + \int_S \gamma(t, y) p(t, y, A) P(s, t; x, dy) \end{aligned}$$

for all $t > s$ (**Kolmogorov's forward equation**).

Further, for every $t \in \mathbb{R}_0^+$, $x \in S$ and $A \in \mathcal{S}$, the transition probability $P(s, t; x, A)$ is differentiable with respect to s , and

$$\begin{aligned} \frac{\partial P(s, t; x, A)}{\partial s} &= - \int_S P(s, t; y, A) q(s; x, dy) \\ &= \gamma(s, x) \left(P(s, t; x, A) - \int_S P(s, t; y, A) p(s, x, dy) \right) \end{aligned}$$

for $s < t$ (**Kolmogorov's backward equation**).

Proof: see Gikhman, Skorokhod [44], p.314 and 316 (corollaries)

These differential equations lead to expressions of the transition probabilities in terms of the infinitesimal transition rates. This can be proven by solving the Kolmogorov differential equations via the method of successive approximations by Picard and Lindelöf. The solutions are stated in the following theorems.

Theorem A.5 The transition probabilities of a Markov jump process are uniquely determined by the infinitesimal rates $q(t; x, A)$. They are computed as

$$P(s, t; x, A) = \sum_{n=0}^{\infty} P_b^{(n)}(s, t; x, A)$$

with

$$P_b^{(0)}(s, t; x, A) := e^{-\int_s^t \gamma(\tau, x) d\tau} 1_A(x)$$

and recursively

$$P_b^{(n+1)}(s, t; x, A) := \int_s^t \int_S P_b^{(n)}(\theta, t; y, A) e^{-\int_s^\theta \gamma(\tau, x) d\tau} \gamma(\theta, x, dy) d\theta$$

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

Proof: see Gikhman, Skorokhod [44], p.318f (corollary and remark)

A. Markov Jump Processes

If one develops this iteration formula, one obtains

$$\begin{aligned}
 P_b^{(n)}(s, t; x, A) = & \underbrace{\int_s^t \int_{u_1}^t \dots \int_{u_{n-1}}^t}_{n \text{ integrals}} \underbrace{\int_S \dots \int_S \int_A}_{n \text{ integrals}} e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x, dx_1) \dots \\
 & \dots e^{-\int_{u_{n-1}}^{u_n} \gamma(u; x_{n-1}) du} \gamma(u_n; x_{n-1}, dx_n) e^{-\int_{u_n}^t \gamma(u; x_n) du} du_n \dots du_1
 \end{aligned} \tag{A.1}$$

as can be shown by induction. The transition probabilities can be computed iteratively by starting with

$$P_0(s, t; x, A) := e^{-\int_s^t \gamma(\tau, x) d\tau} 1_A(x)$$

and iterating

$$P_{n+1}(s, t; x, A) := \int_s^t \int_S P_n(\theta, t; y, A) e^{-\int_s^\theta \gamma(\tau, x) d\tau} \gamma(\theta, x, dy) d\theta + e^{-\int_s^t \gamma(\tau, x) d\tau} 1_A(x)$$

for all $n \in \mathbb{N}_0$ and $s < t \in \mathbb{R}_0^+$. In the limit, this leads to $P(s, t) = \lim_{n \rightarrow \infty} P_n(s, t)$.

The above form of the transition probabilities was derived via the Kolmogorov backward equations and a substitution $P(s, t; x, A) = \exp(-\int_s^t \gamma(\tau, x) d\tau) \cdot g_s(x)$. Another form of the transition probabilities can be derived via the forward equations. They can be solved using the following iteration.

Theorem A.6 *Be X a Markov jump process with infinitesimal rates $q(t; x, A)$. The transition probabilities of X being in some state $y \in A$ after time $t \in \mathbb{R}^+$ under the condition of having been in state x at time $s < t$ are given as*

$$P(s, t; x, A) = \sum_{n=0}^{\infty} P^{(n)}(s, t; x, A)$$

with

$$P^{(0)}(s, t; x, A) := 1_A(x)$$

and recursively

$$P^{(n+1)}(s, t; x, A) := \int_s^t \int_S P^{(n)}(s, u; x, dy) q(u; y, A) du$$

for all $n \in \mathbb{N}_0$ and $s < t \in \mathbb{R}_0^+$.

Proof: By definition, the infinitesimal rates $q(t; x, A)$ are continuous in t for every $x \in S$ and $A \in \mathcal{S}$. Induction by n yields that $P^{(n)}(s, t; x, A)$ is differentiable in t for every $n \in \mathbb{N}_0$.

Now, direct validation shows

$$\begin{aligned}
\frac{\partial}{\partial t} \sum_{n=0}^{\infty} P^{(n)}(s, t; x, A) &= \frac{\partial}{\partial t} \sum_{n=1}^{\infty} P^{(n)}(s, t; x, A) \\
&= \sum_{n=1}^{\infty} \frac{\partial}{\partial t} \int_s^t \int_S P^{(n-1)}(s, u; x, dy) q(u; y, A) du \\
&= \sum_{n=1}^{\infty} \int_S P^{(n-1)}(s, t; x, dy) q(t; y, A) \\
&= \int_S q(t; y, A) \sum_{n=1}^{\infty} P^{(n-1)}(s, t; x, dy)
\end{aligned}$$

i.e. the Kolmogorov forward equation is satisfied. According to Gikhman, Skorokhod [44], p.317 (theorem 3), the solution to the Kolmogorov forward equation is unique.

☺

Again by induction, one can prove that

$$P^{(n)}(s, t; x, A) = \underbrace{\int_s^t \int_s^{u_n} \dots \int_s^{u_2}}_{n \text{ integrals}} (Q(u_1) \dots Q(u_n))(x, A) du_1 \dots du_n \quad (\text{A.2})$$

with $Q(u)$ denoting the generator at time $u \in [s, t]$. An iteration for computing the transition probabilities is given by starting with $P_0(s, t; x, A) := 1_A(x)$ and iterating by

$$P_{n+1}(s, t; x, A) := \int_s^t \int_S P_n(s, u; x, dy) q(u; y, A) du + 1_A(x)$$

for all $n \in \mathbb{N}_0$ and $s < t \in \mathbb{R}_0^+$. In the limit, this leads to $P(s, t) = \lim_{n \rightarrow \infty} P_n(s, t)$. This iteration reflects the method of successive approximation by Picard and Lindelöf for solving the Kolmogorov forward differential equations. For the special case of a finite state space, this formula reduces to the iteration given in Bellman [19], p.168, or Kamke [53], p.52.

In the usual manner, the transition probabilities derived above determine the finite-dimensional marginal probabilities which by Kolmogorov's extension theorem (see Gikhman, Skorokhod [44], p.108) uniquely determine the distribution of the process.

Theorem A.7 *The distribution of a Markov jump process $X = (X(t) : t \in \mathbb{R}_0^+)$ is uniquely determined by its infinitesimal rates $q(t; x, A)$ via its transition probabilities. The finite-dimensional marginal distribution of X at times $t_1 < \dots < t_n \in \mathbb{R}_0^+$ with $n \in \mathbb{N}$ is given by*

$$P(X(t_1) \in A_1, \dots, X(t_n) \in A_n) =$$

$$\int_S \pi(dx_0) \int_{A_1} P(t_0, t_1; x_0, dx_1) \dots \int_{A_{n-1}} P(t_{n-2}, t_{n-1}; x_{n-2}, dx_{n-1}) P(t_{n-1}, t_n; x_{n-1}, A_n)$$

if π denotes an initial distribution at time $t_0 < t_1$.

A. Markov Jump Processes

Proof: see Gikhman, Skorokhod [44], p.142

Thus definition A.1 uniquely determines a stochastic process. Its finite-dimensional marginal distributions can be computed knowing the initial distribution at a time t_0 and the infinitesimal transition rates.

For the case of an homogeneous Markov jump process, the infinitesimal transition rates are constant in time, and we can define

$$q(x, A) := q(t; x, A)$$

for all $t \in \mathbb{R}_0^+$, $x \in S$ and $A \in \mathcal{S}$. Further, the conditional probabilities $p(t; x, A)$ from definition A.2 are constant in t , and we can write

$$p(x, A) := p(t; x, A)$$

for all $t \in \mathbb{R}_0^+$, $x \in S$ and $A \in \mathcal{S}$. Further define the kernels Q and P by $Q(x, A) := q(x, A)$ and $P(x, A) := p(x, A)$ for all $x \in S$ and $A \in \mathcal{S}$. The kernel Q shall be called the generator of X . The kernel P is called the jump kernel of X .

Then the transition probability kernel assumes a particularly tractable form:

Theorem A.8 *Let X denote an homogeneous Markov jump process with infinitesimal transition rates ($q(x, A) : x \in S, A \in \mathcal{S}$). Then the transition probability kernel $P(s, t)$ can be written as*

$$P(s, t) = e^{Q(t-s)} := \sum_{k=0}^{\infty} \frac{(t-s)^k}{k!} Q^k$$

for all $s < t$.

Proof: It suffices to prove

$$P^{(k)}(s, t) = \frac{(t-s)^k}{k!} Q^k$$

for all $k \in \mathbb{N}_0$ and $s < t$. This is done by induction on k . Beginning with the obvious equality $P^{(0)}(s, t) = I = \frac{(t-s)^0}{0!} Q^0$, the iteration step of theorem A.6 leads to

$$\begin{aligned} P^{(k+1)}(s, t; x, A) &= \int_s^t \int_S P^{(k)}(s, u; x, dy) q(u; y, A) du \\ &= \int_s^t \int_S \frac{(u-s)^k}{k!} Q^k(x, dy) q(y, A) du \end{aligned}$$

for all $k \in \mathbb{N}_0$ by induction hypothesis and homogeneity of the transition rates. This yields further

$$\begin{aligned} P^{(k+1)}(s, t; x, A) &= \int_s^t \frac{(u-s)^k}{k!} du \int_S Q^k(x, dy) q(y, A) \\ &= \frac{(t-s)^{k+1}}{(k+1)!} Q^{k+1}(x, A) \end{aligned}$$

which proves the induction step.

☺

Remark A.5 In the homogeneous case, the computation of the transition probabilities can be accomplished by solving the Kolmogorov forward differential equations

$$\frac{\partial}{\partial t}P(s, t) = P(s, t)Q$$

with initial value $P(s, s) = I$ via the Runge–Kutta method. This is possible in a straightforward way without encountering numerical problems (see Asmussen [6], Gilbert [43] or Moler, van Loan [67]).

In some cases the computation of the transition probabilities of general Markov jump processes can be reduced to computing some combination of transition probability kernels of homogeneous Markov jump processes. This shall be described in the following.

Choose some time interval $]s, t]$ over which to compute the transition probability kernel. Assume that there is a step function $\hat{Q} : t \rightarrow \hat{Q}(t)$, with $\hat{Q}(t)$ being a generator for every t , which approximates the function $Q : t \rightarrow Q(t)$ uniformly in $]s, t]$, i.e.

$$\|\hat{Q} - Q\|_{s,t} := \sup_{s \leq u \leq t} \|\hat{Q}(u) - Q(u)\| = \sup_{s \leq u \leq t, x \in S, A \in \mathcal{S}} |\hat{Q}(u)(x, A) - Q(u)(x, A)| < \varepsilon$$

for some $\varepsilon > 0$. Then the following approximation holds:

Theorem A.9 *Let $\hat{Q} : u \rightarrow \hat{Q}(u)$ be a step function such that $\|\hat{Q} - Q\|_{s,t} < \varepsilon$ and $\hat{Q}(u)$ is a generator on (S, \mathcal{S}) for every $u \in [s, t]$. Then the transition probability kernel $P(s, t)$ can be approximated by*

$$\hat{P}(s, t) = \prod_{i=1}^n e^{\hat{Q}_i \cdot (s_i - s_{i-1})}$$

for $s_0 := s$, $s_n := t$ and \hat{Q} being defined by $\hat{Q}(u) = \hat{Q}_i$ on $u \in]s_{i-1}, s_i]$ for $i \in \{1, \dots, n\}$. The error of this approximation is bounded by

$$\|\hat{P}(s, t) - P(s, t)\| < \varepsilon \cdot (t - s)$$

A. Markov Jump Processes

Proof: Write $\hat{Q}(u) = Q(u) + \Delta(u)$ for all $u \in [s, t]$. Then by assumption $\|\Delta(u)\| < \varepsilon$ for all $u \in [s, t]$. Now theorem A.6, formula A.2 and theorem A.8 yield

$$\begin{aligned}
& \hat{P}(s, t) - P(s, t) = \\
&= \sum_{k=0}^{\infty} \underbrace{\int_s^t \int_s^{u_k} \dots \int_s^{u_2}}_{k \text{ integrals}} (Q(u_1) + \Delta(u_1)) \dots (Q(u_k) + \Delta(u_k)) du_1 \dots du_k \\
&\quad - \sum_{k=0}^{\infty} \underbrace{\int_s^t \int_s^{u_k} \dots \int_s^{u_2}}_{k \text{ integrals}} Q(u_1) \dots Q(u_k) du_1 \dots du_k \\
&= \sum_{k=1}^{\infty} \underbrace{\int_s^t \int_s^{u_k} \dots \int_s^{u_2}}_{k \text{ integrals}} \hat{Q}(u_1) \dots \hat{Q}(u_{k-1}) \Delta(u_k) du_1 \dots du_k \\
&\quad + \sum_{k=2}^{\infty} \underbrace{\int_s^t \int_s^{u_k} \dots \int_s^{u_2}}_{k \text{ integrals}} \hat{Q}(u_1) \dots \hat{Q}(u_{k-2}) \Delta(u_{k-1}) Q(u_k) du_1 \dots du_k \\
&\quad + \sum_{k=3}^{\infty} \underbrace{\int_s^t \int_s^{u_k} \dots \int_s^{u_2}}_{k \text{ integrals}} \hat{Q}(u_1) \dots \hat{Q}(u_{k-3}) \Delta(u_{k-2}) Q(u_{k-1}) Q(u_k) du_1 \dots du_k \\
&\quad + \dots \\
&= \sum_{k=0}^{\infty} \int_s^t \hat{P}(s, u) \Delta(u) \underbrace{\int_u^t \int_u^{u_k} \dots \int_u^{u_2}}_{k \text{ integrals}} Q(u_1) \dots Q(u_k) du_1 \dots du_k du
\end{aligned}$$

by a similar calculation as in the end of the proof of theorem 2.8. Hence we have

$$\hat{P}(s, t) - P(s, t) = \int_s^t \hat{P}(s, u) \Delta(u) P(u, t) du$$

and because of $\|\hat{P}(s, u)\| = \|P(u, t)\| = 1$ for all $u \in [s, t]$, we get

$$\|\hat{P}(s, t) - P(s, t)\| < \varepsilon \cdot (t - s)$$

which proves the approximation bound.

☺

Remark A.6 Since the function $Q : t \rightarrow Q(t)$ is continuous with respect to the supremum norm (see remark A.1), the existence of an approximating step function \hat{Q} as assumed in theorem A.9 is immediate. The function \hat{Q} defines a piecewise homogeneous Markov jump process.

In the rest of this section, the main properties of the sample paths are derived. It turns out that the paths behave in an intuitively lucid way: Starting in some state at an initial time, a path remains in that state for a positive holding time, for which the distribution can be determined. Then it jumps to another state according to the probability law given by the jump kernel at the jump time. Remaining in this state for a positive duration, the path jumps again, and so on.

In order to examine path properties, it is necessary to obtain versions of the process which have certain features of their sample paths that make them easier to analyze. This does not pose a real problem, since it is possible to choose those versions such that the probability law of the process is not altered. Hence all the path properties derived are valid almost surely with respect to the distribution of the process (i.e. for all finite-dimensional marginal distributions, according to Kolmogorov's extension theorem).

Definition A.3 Two stochastic processes $X = (X_t : t \in \mathbb{R}_0^+)$ and $Y = (Y_t : t \in \mathbb{R}_0^+)$ are called **stochastically equivalent** if their finite-dimensional marginal distributions coincide, i.e. if for every $n \in \mathbb{N}$ and t_1, \dots, t_n the distributions

$$P(X(t_1), \dots, X(t_n)) = P(Y(t_1), \dots, Y(t_n))$$

are equal. If X and Y are equivalent, then Y is called a **version** of X .

In order to further examine the properties of Markov jump process, it is necessary to introduce the concept of separability. For separable processes, limits over continuous sets (such as $\inf\{t \in \mathbb{R}_0^+ : X(t) \in A\}$) are still measurable and path properties such as first hitting times can be explored.

Definition A.4 A stochastic process $X = ((X(t) : t \in \mathbb{R}_0^+)$ with state space S is called **separable** if there is a countable and everywhere dense subset S of \mathbb{R}_0^+ and a negligible set $N \subset \Omega$ such that for every open set $G \subset \mathbb{R}_0^+$ and every closed set $F \subset S$ the set difference

$$\{X(t) \in F \forall t \in G \cap S\} \setminus \{X(t) \in F \forall t \in G\} \subset N$$

is negligible. The set S is called **set of separability**.

Definition A.5 Let $X = ((X(t) : t \in \mathbb{R}_0^+)$ be a stochastic process with metric state space S . Denote the metric in S by ρ . The process X is called **stochastically continuous** if for every time $t \in \mathbb{R}_0^+$ and every $\varepsilon > 0$ the limit

$$\lim_{h \rightarrow 0} P(\rho(X(t), X(t+h)) > \varepsilon) = 0$$

holds.

Theorem A.10 *A Markov jump process is stochastically continuous.*

Proof: see Gikhman, Skorokhod [44], p.347 (lemma 1)

A. Markov Jump Processes

Theorem A.11 *For a Markov jump process $X = ((X(t) : t \in \mathbb{R}_0^+)$ with a separable and locally compact metric state space S and for every countable and everywhere dense subset $S \subset \mathbb{R}_0^+$, there is an equivalent process $Y = ((Y(t) : t \in \mathbb{R}_0^+)$ which is separable and has the set of separability S .*

Proof: see Gikhman, Skorokhod [44], p.153 (theorem 2), p.155 (theorem 5) and the above theorem A.10

Now the first result for the sample path behaviour can be derived. In the following, we always examine the separable version of a Markov jump process. A natural generalization of the distribution of holding times for homogeneous Markov jump processes is given by

Theorem A.12 *For a separable version X of a Markov jump process, the holding time distribution in a state $x \in S$ is*

$$P(X(\tau) = x \forall \tau \in [s, t] | X(s) = x) = e^{-\int_s^t \gamma(\tau, x) d\tau}$$

for every $s < t$.

Proof: see Gikhman, Skorokhod [44], p.348 (lemma 2)

For further path properties, one needs to require right continuity of the paths, which can be done without altering the probability law of the process:

Theorem A.13 *For a separable Markov jump process there exists an equivalent process with sample paths that are continuous from the right with probability 1.*

Proof: see Gikhman, Skorokhod [44], p.349 (corollary 1)

Theorem A.14 *If X is a separable Markov jump process and $X(t) = x$, then with probability 1 there exists an interval of time $[t, t + h[$ during which $X(t) = x$ for all $\tau \in [t, t + h[$.*

Proof: see Gikhman, Skorokhod [44], p.349 (corollary 2)

Assume in the following that the Markov jump process X is a separable version which has only paths that are continuous from the right. Then the next theorems yield a more accurate description of its sample path behaviour than the finite-dimensional marginal distributions can give:

Theorem A.15 *Denote the times between jumps of X by $(\tau_n : n \in \mathbb{N})$ starting with the first jump after a time instant $t_0 \in \mathbb{R}_0^+$. Be $n \in \mathbb{N}$, $t_1, \dots, t_n \in \mathbb{R}_0^+$ and $A_1, \dots, A_n \in \mathcal{S}$. Let B denote the event*

$$B = \left\{ \forall k \in \{1, \dots, n\} : \forall t \in \left[t_0 + \sum_{j=1}^{k-1} \tau_j, t_0 + \sum_{j=1}^k \tau_j \right[: X(t) = \xi_{k-1} \right. \\ \left. \wedge X \left(t_0 + \sum_{j=1}^n \tau_j \right) = \xi_n \wedge \xi_{k-1} \neq \xi_k \wedge \xi_k \in A_k \wedge \tau_k < t_k \right\}$$

Then the conditional probability of B under the hypothesis that $X(t_0) = x_0$ is given by

$$\begin{aligned} F(t_1, \dots, t_n; A_1, \dots, A_n | t_0, x_0) = \\ = \int_{t_0}^{t_0+t_1} e^{-\int_{t_0}^{s_1} \gamma(\tau, x_0) d\tau} d s_1 \int_{A_1} \gamma(s_1, x_0, dx_1) \int_{s_1}^{s_1+t_2} e^{-\int_{s_1}^{s_2} \gamma(\tau, x_1) d\tau} d s_2 \dots \\ \dots \int_{s_{n-1}}^{s_{n-1}+t_n} e^{-\int_{s_{n-1}}^{s_n} \gamma(\tau, x_{n-1}) d\tau} d s_n \gamma(s_n, x_{n-1}, A_n) \end{aligned}$$

Proof: see Gikhman, Skorokhod [44], p.351f (formula 8)

Theorem A.16 *With probability 1, the Markov jump process X has only finitely many jumps in any finite interval of time.*

Proof: see Gikhman, Skorokhod [44], p.353 (corollary 1)

Theorem A.17 *The sample paths of the separable version of a Markov jump process that is continuous from the right have with probability 1 the following structure: There is a sequence of positive random variables $(\tau_k : k \in \mathbb{N})$ with*

$$\lim_{m \rightarrow \infty} \sum_{k=1}^m \tau_k = \infty$$

and for every time t with

$$\sum_{k=1}^m \tau_k \leq t < \sum_{k=1}^{m+1} \tau_k$$

the state at time t is $X(t) = X(\tau_m)$.

Proof: see Gikhman, Skorokhod [44], p.354 (corollary 2)

Remark A.7 The above theorem shows the probabilistic meaning of the infinitesimal rates. Starting in some state $x \in S$ at an initial time $t_0 := 0$, almost every path remains in this state for a time duration which is distributed according to theorem A.12. Upon a time instant $\tau \in \mathbb{R}_0^+$ at which the state x is left, some new state $y \in A$ is assumed according to the probability $p(\tau, x, A)$. Then the path behaves in the same way again, starting in state y and with initial time τ .

The following theorem proves the intuitive meaning of the components $P_b^{(n)}(s, t)$ appearing in the form A.1 of the transition probabilities.

A. Markov Jump Processes

Theorem A.18 Let X be a Markov jump process with measurable state space (S, \mathcal{S}) . Denote the times between jumps of X by $(\tau_n : n \in \mathbb{N})$ starting with the first jump after a time instant $s \in \mathbb{R}_0^+$. Be $n \in \mathbb{N}_0$, $s < t \in \mathbb{R}^+$ and $A_1, \dots, A_n \in \mathcal{S}$. Let $C(n, s, t)$ denote the event

$$C(n, s, t) = \left\{ \forall t \in [s, s + \tau_1[: X(t) = x \right. \\ \wedge \forall k \in \{1, \dots, n-1\} : \forall t \in \left[s + \sum_{j=1}^k \tau_j, s + \sum_{j=1}^{k+1} \tau_j \right[: X(t) \in A_k \\ \left. \wedge \forall t \in \left[s + \sum_{j=1}^n \tau_j, t \right] : X(t) \in A_n \right\}$$

Then the conditional probability of $C(n, s, t)$ under the hypothesis that $X(s) = x$ is given by

$$P_x^{(n)}(s, t; A_1, \dots, A_n) = \underbrace{\int_s^t \int_{u_1}^t \dots \int_{u_{n-1}}^t}_{n \text{ integrals}} \underbrace{\int_{A_1} \dots \int_{A_n}}_{n \text{ integrals}} e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x, dx_1) \dots \\ \dots e^{-\int_{u_{n-1}}^{u_n} \gamma(u; x_{n-1}) du} \gamma(u_n; x_{n-1}, dx_n) e^{-\int_{u_n}^t \gamma(u; x_n) du} du_n \dots du_1$$

Proof: This is shown by induction on n . For $n = 1$, the event $C(n, s, t)$ means that there is exactly one jump during $]s, t]$ and at this instant, X changes from state x to some state $x_1 \in A_1$. By theorem A.12, the density of the holding time in a state $x \in S$ with respect to the Lebesgue measure on the time axis is

$$\lim_{h \rightarrow 0} \frac{P_x(T \in [t, t+h])}{h} = \lim_{h \rightarrow 0} \frac{e^{-\int_s^t \gamma(u; x) du} - e^{-\int_s^{t+h} \gamma(u; x) du}}{h} = e^{-\int_s^t \gamma(u; x) du} \cdot \gamma(t; x)$$

since $\gamma(\cdot; x) : t \rightarrow \gamma(t; x)$ is continuous by definition. Under the condition that the holding time in x is from s to u_1 , the event $C(n, s, t)$ occurs if and only if the holding time in the new state x_1 is at least from u_1 to t . Remembering definition A.2 and conditioning on the holding time in the initial state x leads to

$$P_x^{(1)}(s, t; A_1) = \int_s^t e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x) \cdot \int_{A_1} p(u_1; x, dx_1) \cdot e^{-\int_{u_1}^t \gamma(u; x_1) du} du_1 \\ = \int_s^t \int_{A_1} e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x, dx_1) e^{-\int_{u_1}^t \gamma(u; x_1) du} du_1$$

Now assume that the statement is true for $n \in \mathbb{N}$. The event $C(n+1, s, t)$ means that there is exactly one jump from the initial state x to some state $x_1 \in A_1$ at a time $u_1 \in]s, t]$ and after this the event $C(n, u_1, t)$ occurs under initial state x_1 . Conditioning on the holding time in the

initial state x and using the induction hypothesis leads to

$$\begin{aligned}
P_x^{(n+1)}(s, t; A_1, \dots, A_{n+1}) &= \\
&= \int_s^t e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x) \cdot \int_{A_1} p(u_1; x, dx_1) \cdot P_{x_1}^{(n)}(u_1, t; A_2, \dots, A_{n+1}) du_1 \\
&= \int_s^t \int_{A_1} e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x, dx_1) \underbrace{\int_{u_1}^t \dots \int_{u_n}^t}_{n \text{ integrals}} \underbrace{\int_{A_2} \dots \int_{A_{n+1}}}_{n \text{ integrals}} e^{-\int_{u_1}^{u_2} \gamma(u; x) du} \gamma(u_2; x_1, dx_2) \dots \\
&\quad \dots e^{-\int_{u_n}^{u_{n+1}} \gamma(u; x_n) du} \gamma(u_{n+1}; x_n, dx_{n+1}) e^{-\int_{u_{n+1}}^t \gamma(u; x_{n+1}) du} du_{n+1} \dots du_2 du_1
\end{aligned}$$

which completes the induction step.

☺

Definition A.6 Let X be a Markov jump process with measurable state space (S, \mathcal{S}) and $A \in \mathcal{S}$ some measurable subset. Then the **first hitting time** of A is defined by

$$T_A := \inf\{t \in \mathbb{R}^+ : X(t) \in A\}$$

Theorem A.19 The first hitting time T_A is measurable and has the distribution

$$\begin{aligned}
P_x(T_A > t, X_t \in B) &= \sum_{n=0}^{\infty} \underbrace{\int_0^t \int_{u_1}^t \dots \int_{u_{n-1}}^t}_{n \text{ integrals}} \underbrace{\int_{A^c} \dots \int_{A^c} \int_B}_{n \text{ integrals}} e^{-\int_0^{u_1} \gamma(u; x) du} \gamma(u_1; x, dx_1) \dots \\
&\quad \dots e^{-\int_{u_{n-1}}^{u_n} \gamma(u; x_{n-1}) du} \gamma(u_n; x_{n-1}, dx_n) e^{-\int_{u_n}^t \gamma(u; x_n) du} du_n \dots du_1
\end{aligned}$$

for every $t \in \mathbb{R}^+$, initial state $x \in A^c$ and final set $B \subset A$, $B \in \mathcal{S}$.

Proof: T_A is measurable, because X is separable (theorem A.11). The form of the distribution follows from theorem A.18 by summing up over the number n of jumps in $]0, t]$.

☺

As already mentioned in remark A.3, a Markov jump process usually cannot be constructed by specifying the transition probabilities or finite-dimensional marginal distributions, since these are not available. Even more, these belong to the values one wishes to derive from the stochastic model. The most intuitive values that one can start from are the tendencies of a Markov jump process at a certain time and in a certain state, i.e. the infinitesimal transition rates. The rest of this section shortly describes how to do that.

Definition A.7 Any family $(q(t; x, A) : t \in \mathbb{R}_0^+, x \in S, A \in \mathcal{S})$ of real numbers that satisfies

$$q(t; x, A) = -\gamma(t, x)1_A(x) + \gamma(t, x, A)$$

for all $t \in \mathbb{R}_0^+$, $x \in S$, $A \in \mathcal{S}$, and further the conditions

A. Markov Jump Processes

1. For fixed $t \in \mathbb{R}_0^+$ and $x \in S$, the function $\gamma(t, x, A)$ on S is a finite measure and $\gamma(t, x) = \gamma(t, x, S)$.
2. For fixed $(x, A) \in S \times \mathcal{S}$ the function $\gamma(t, x, A)$ is continuous in t .
3. For fixed $x \in S$, the measure $\gamma(t, x, \cdot)$ is continuous in t (with respect to the vague topology).
4. The values $\gamma(t, x)$ are bounded in $(t, x) \in \mathbb{R}_0^+ \times S$.

is called a family of **regular rates** .

Remark A.8 Denote the space of generators on S by \mathcal{G} . Using the same norm on the space of kernels on S as introduced in remark A.1, the above conditions can be summarized as follows: A bounded continuous function $Q : \mathbb{R}_0^+ \rightarrow \mathcal{G}$ is called a **regular generator function**. A regular generator function provides a family of regular rates by $q(t; x, A) := Q(t)(x, A)$.

Consider the space of sequences $\Omega = \{\omega = (\xi_n, \tau_n : n \in \mathbb{N}) : \xi_k \in A_k, \tau_k \in \mathbb{R}_0^+\}$. On the algebra of cylindrical sets in Ω introduce the measure P as follows. Using the distribution F from theorem A.15, define

$$P(C) := \int_S F(t_1, \dots, t_n; A_1, \dots, A_n | 0, x) \mu(dx)$$

for every set

$$C = \{\xi_1 \in A_1, \dots, \xi_n \in A_n; \tau_1 \leq t_1, \dots, \tau_n \leq t_n\}$$

and any initial distribution μ on S which is assumed at time 0.

According to Kolmogorov's extension theorem (see Gikhman, Skorokhod [44], p.108), the measure P can be extended to the σ -algebra \mathcal{F} generated by the cylindrical sets in Ω . This defines a probability space (Ω, \mathcal{F}, P) . Now define a random function $X = (X(t) : t \in \mathbb{R}_0^+)$ by

$$X(t, \omega) := \xi_k(\omega)$$

for all

$$\sum_{j=1}^k \tau_j(\omega) \leq t < \sum_{j=1}^{k+1} \tau_j(\omega)$$

Then X has the following properties:

Theorem A.20 *The process X constructed above is a Markov process and has almost certainly sample paths which are continuous from the right. The transition probabilities of X are given by*

$$P(s, t; x, A) = \sum_{n=0}^{\infty} P_b^{(n)}(s, t; x, A) \quad (\text{A.3})$$

with

$$P_b^{(0)}(s, t; x, A) := e^{-\int_s^t \gamma(\tau, x) d\tau} 1_A(x)$$

and

$$P_b^{(n+1)}(s, t; x, A) := \int_s^t \int_S P_b^{(n)}(\theta, t; y, A) e^{-\int_s^\theta \gamma(\tau, x) d\tau} \gamma(\theta, x, dy) d\theta$$

for all $s < t \in \mathbb{R}_0^+$, $x \in S$ and $A \in \mathcal{S}$. They satisfy the Kolmogorov backward equation

$$\frac{\partial P(s, t; x, A)}{\partial s} = - \int_S P(s, t; y, A) q(t; x, dy)$$

with boundary condition

$$\lim_{s \uparrow t} P(s, t; x, A) = 1_A(x)$$

Finally, the equation

$$\lim_{h \downarrow 0} \frac{P(t, t+h; x, A) - 1_A(x)}{h} = q(t; x, A)$$

holds uniformly with respect to t (for $0 \leq t \leq T$ with arbitrary T) for fixed $x \in S$ and $A \in \mathcal{S}$.

Proof: see Gikhman, Skorokhod [44], p.364 (theorem 4) and p.363

Hence the process constructed in the described way is a Markov jump process with the desired infinitesimal transition rates.

A.2. Quasi-Commutability of Generators

A special case of Markov jump processes is given, if the following equation A.4 is satisfied for the generators. This condition states that the differentiation rule for a function of the generators can be applied as if the generators were commutable. Therefore the condition shall be called quasi-commutability of the generators.

As in the homogeneous case, the transition probabilities as well as the transient distribution assume a rather simple form, which is derived first. After that, two special cases are examined for which this simple form leads to the computation of asymptotic distributions.

A. Markov Jump Processes

Definition A.8 If the equation

$$\frac{d}{dt} \left(\int_s^t Q(u) du \right)^k = k \cdot \left(\int_s^t Q(u) du \right)^{k-1} Q(t) \quad (\text{A.4})$$

holds for all $k \in \mathbb{N}$ and $s < t$, the generators $(Q(t) : t \in \mathbb{R}_0^+)$ shall be called **quasi-commutable**.

Remark A.9 A sufficient condition for equation A.4 is given if for every $s < t \in \mathbb{R}_0^+$ the kernels $Q(t)$ and $\int_s^t Q(u) du$ are commutable. A sufficient condition for this is the special form

$$Q(t) = \lambda(t) \cdot Q$$

for the generators $Q(t)$ of X , with $\lambda : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ describing some time-dependent intensity and $Q : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ denoting any generator.

Theorem A.21 If the generators of X are quasi-commutable, then the transition probabilities $P(s, t)$ assume the form

$$P(s, t) = e^{\int_s^t Q(u) du}$$

Proof: For every $k \in \mathbb{N}$ it will be shown that

$$\underbrace{\int_s^t \int_s^{u_k} \dots \int_s^{u_2}}_{k \text{ integrals}} Q(u_1) \dots Q(u_k) du_1 \dots du_k = \frac{\left(\int_s^t Q(u) du \right)^k}{k!}$$

is implied by equation A.4. For $k = 0$ this holds by definition and for $k = 1$ this is obvious. The induction step for $k+1$ is seen by first applying the induction hypothesis and then equation A.4 in

$$\begin{aligned} \underbrace{\int_s^t \int_s^{u_{k+1}} \dots \int_s^{u_2}}_{k+1 \text{ integrals}} Q(u_1) \dots Q(u_{k+1}) du_1 \dots du_{k+1} &= \int_s^t \frac{\left(\int_s^{u_{k+1}} Q(u) du \right)^k}{k!} Q(u_{k+1}) du_{k+1} \\ &= \frac{1}{k!} \int_s^t \frac{1}{k+1} \frac{d}{du_{k+1}} \left(\int_s^{u_{k+1}} Q(u) du \right)^{k+1} du_{k+1} = \frac{1}{(k+1)!} \left(\int_s^t Q(u) du \right)^{k+1} \end{aligned}$$

Now the statement follows by theorem A.6.

☺

Example A.1 Assume that there is a function $\lambda : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ such that $Q(t) = \lambda(t) \cdot Q$ for all $t \in \mathbb{R}_0^+$. Then the transition probability kernel from time $s \in \mathbb{R}_0^+$ to time $t > s$ can be simplified to the form

$$P(s, t) = e^{\int_s^t \lambda(u) du \cdot Q} = \sum_{k=0}^{\infty} \frac{\left(\int_s^t \lambda(u) du \right)^k}{k!} Q^k$$

Definition A.9 For a stochastic process $X = (X_t : t \in \mathbb{R}_0^+)$ with initial distribution π at time $t_0 := 0$ and transition probabilities $P(s, t; x, A)$, the **transient distribution** at time $t \in \mathbb{R}_0^+$ is defined as

$$X_t^\pi := P_\pi^{X_t} = \int_S d\pi P(0, t)$$

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

Theorem A.22 *If condition A.4 is satisfied, then the transient distribution under initial distribution π assumes the form*

$$X_t^\pi = \int d\pi \sum_{k=0}^{\infty} \frac{\left(\int_0^t Q(u)du\right)^k}{k!} = \int d\pi e^{\int_0^t Q(u)du}$$

Proof: This is immediate by theorem A.21.

Concluding this section, the results for asymptotic distributions in special cases are shown, after first some estimations need to be proven.

Definition A.10 Define the norm of **total variation** by

$$\|\nu\| := |\nu|(S) := \int_S d|\nu|$$

for all charges ν on \mathcal{S} .

Definition A.11 Using the norm of total variation for charges, define the **supremum norm for kernels** by

$$\|K\| := \sup_{x \in S} \|K(x, \cdot)\| = \sup_{x \in S} |K(x, \cdot)|(S)$$

for all kernels K on $S \times S$.

Theorem A.23 *Let K denote a kernel on $S \times S$ with $\|K\| < \varepsilon$. Then for every probability measure μ on S , the estimation*

$$\|\mu(e^K - I)\| < 2\varepsilon$$

holds with I denoting the identity kernel.

Proof: For $n = 1$, the inequality $\|\mu K^n\| < \varepsilon^n$ is valid, since μ is a probability measure. Assume that this inequality holds for some $n \in \mathbb{N}$. Then

$$\|\mu K^{n+1}\| = \left\| \int_S d\mu(x) K^{n+1}(x, \cdot) \right\| = \left\| \int_S d(\mu K^n)(x) K(x, \cdot) \right\| \leq \varepsilon \cdot \|\mu K^n\| < \varepsilon^{n+1}$$

A. Markov Jump Processes

completes the induction step. Hence,

$$\|\mu(e^K - I)\| = \left\| \mu \sum_{n=1}^{\infty} \frac{K^n}{n!} \right\| \leq \sum_{n=1}^{\infty} \frac{\|\mu K^n\|}{n!} < e^\varepsilon - 1 < 2\varepsilon$$

☺

Theorem A.24 *Let X denote a Markov jump process with generator $Q(t)$ at time $t \in \mathbb{R}_0^+$. Assume that the generator converges in the following strong sense: There is a generator Q with $\|Q\| > 0$ such that for every $\varepsilon > 0$ there is a time $t_0 \in \mathbb{R}^+$ with*

$$\left\| \int_{t_0}^t Q(u) du - Q \cdot (t - t_0) \right\| < \varepsilon$$

for all $t \geq t_0$. Further assume that the homogeneous Markov jump process with generator Q is positive recurrent with stationary distribution q . Then q is the asymptotic distribution for X , i.e. for every initial distribution π ,

$$\|X_t^\pi - q\| \rightarrow 0 \text{ as } t \rightarrow \infty$$

Proof: Choose any $\varepsilon > 0$ and $t_0 \in \mathbb{R}^+$ such that $\left\| \int_{t_0}^t Q(u) du - (t - t_0)Q \right\| < \varepsilon$ for all $t \geq t_0$. Since Q is positive recurrent, there is a time $t_1 \in \mathbb{R}^+$ such that $\|\pi e^{Q t} - q\| < \varepsilon$ for all $t \geq t_1$. Set $T := \max\{t_0, t_1\}$. Then for $t > T$ and any initial distribution π of X ,

$$\begin{aligned} \|X_t^\pi - q\| &\leq \left\| X_T^\pi e^{\int_T^t Q(u) du} - X_T^\pi e^{Q \cdot (t - T)} \right\| + \left\| X_T^\pi e^{Q \cdot (t - T)} - q \right\| \\ &< \left\| X_T^\pi \left(e^{\int_T^t Q(u) du} - e^{Q \cdot (t - T)} \right) \right\| + \varepsilon \end{aligned}$$

Abbreviating $A := \int_T^t Q(u) du$ and $B := Q \cdot (t - T)$, the remaining term can be estimated via the argument

$$\begin{aligned} \|e^A - e^B\| &= \left\| \sum_{k=1}^{\infty} \frac{A^k - B^k}{k!} \right\| = \left\| \sum_{k=1}^{\infty} \frac{((A - B) + B)^k - B^k}{k!} \right\| \\ &\leq \left\| \sum_{k=1}^{\infty} \frac{(A - B)^k}{k!} \right\| + \left\| (A - B) \sum_{k=1}^{\infty} \frac{\sum_{l=1}^k \binom{k}{l} (A - B)^{l-1} B^{k-l}}{k!} \right\| \\ &< (e^\varepsilon - 1) + \varepsilon \cdot \left\| \sum_{l=1}^{\infty} \sum_{k=l}^{\infty} \frac{(A - B)^{l-1} B^{k-l}}{l!(k-l)!} \right\| \\ &< (e^\varepsilon - 1) + \varepsilon \cdot \sum_{l=1}^{\infty} \frac{\|A - B\|^{l-1}}{(l-1)!} \cdot \sum_{k=0}^{\infty} \frac{\|B\|^k}{k!} \\ &\leq (e^\varepsilon - 1) + \varepsilon \cdot e^\varepsilon \cdot 1 < 4\varepsilon \end{aligned}$$

Using this, the estimation

$$\|X_t^\pi - q\| < 5\varepsilon$$

holds for all $t > T$, since X_T^π is a probability distribution.

☺

Example A.2 Assume that the generator function has the form $Q(t) = Q + \frac{1}{t^2+1}R$ for all $t \in \mathbb{R}_0^+$, with constant generators Q and R . Further assume that the homogeneous Markov jump process with generator Q is positive recurrent with stationary distribution q . Then the above theorem applies (choose $t_0 := \frac{2}{\varepsilon \cdot \|R\|}$) and the Markov jump process with generator $Q(t)$ has asymptotic distribution q .

Theorem A.25 Let X denote a Markov jump process with generator $Q(t)$ at time $t \in \mathbb{R}_0^+$. If the integral

$$Q := \int_0^\infty Q(u)du$$

yields a kernel with $\|Q\| < \infty$, then the asymptotic distribution of X under initial distribution π is given by

$$q = \int_S d\pi e^{\int_0^\infty Q(u)du}$$

Proof: Choose any $\varepsilon > 0$. The assumption $\|\int_0^\infty Q(u)du\| < \infty$ implies that there is a time $t_0 \in \mathbb{R}^+$ such that $\|\int_t^\infty Q(u)du\| < \varepsilon$ for all $t \geq t_0$. Then

$$\|X_t^\pi - q\| = \left\| \pi e^{\int_0^t Q(u)du} - \pi e^{\int_0^\infty Q(u)du} \right\| = \left\| \pi e^{\int_0^t Q(u)du} \left(I - e^{\int_t^\infty Q(u)du} \right) \right\|$$

holds for all $t \in \mathbb{R}^+$. By theorem A.23, the estimation $\|I - e^{\int_t^\infty Q(u)du}\| < 2\varepsilon$ is valid for all $t > t_0$. Since $X_t^\pi = \pi e^{\int_0^t Q(u)du}$ is a probability measure for all $t \in \mathbb{R}^+$, it follows that

$$\|X_t^\pi - q\| = \left\| X_t^\pi \left(I - e^{\int_t^\infty Q(u)du} \right) \right\| < 2\varepsilon$$

for all $t > t_0$.

☺

Remark A.10 Note that this asymptotic distribution is not independent of the initial distribution π . The above theorem represents the pathological case of all states becoming absorbing as the time proceeds. Thus every path of the process reaches an absorbing state at some time. For illustration, one can say that the process falls asleep.

Example A.3 Let R denote the generator of a positive recurrent homogeneous Markov jump process. For the generator function $Q(t) = \lambda \cdot e^{-\lambda t}R$ with $\lambda > 0$, we have $\int_0^\infty Q(u)du = R$. Hence the Markov jump process with generator $Q(t)$ has the same asymptotic distribution as the homogeneous Markov jump process with generator R .

A.3. Periodic Markov Jump Processes

If the infinitesimal transition rates of a Markov jump process are periodic, the Markov jump process will be called periodic. The transient and asymptotic distributions for this special case can be determined by examining the imbedded homogeneous Markov chain at times which are multiples of the period length.

For the transient distribution one can prove a recursion formula which simplifies its computation by reducing the time range of integration in formula A.2 to at most the period length. An asymptotic distribution does exist if the imbedded Markov chain is positive recurrent. Furthermore, conditions for aperiodicity and uniform recurrence of the imbedded Markov chain are derived.

Definition A.12 A Markov jump process X with generator Q is called **periodic** with **period** $T > 0$ if $Q(s) = Q(s + T)$ for all $s \in [0, T[$.

The computation of the transient distribution can be simplified as follows. The periodicity of the generator yields

$$P(0, nT) = P(0, (n-1)T)P((n-1)T, nT) = P(0, (n-1)T)P(0, T) = P(0, T)^n$$

Let π denote the initial distribution of the queue process Q . Define

$$\lfloor t/T \rfloor := \max\{n \in \mathbb{N}_0 : nT \leq t\}$$

as the number of period lengths that have passed until time $t \in \mathbb{R}^+$. Now the transient distribution of Q is given by

$$X_t^\pi = \int d\pi P(0, \lfloor t/T \rfloor T) P(\lfloor t/T \rfloor T, t) = \int d\pi P(0, T)^{\lfloor t/T \rfloor} P(0, t - \lfloor t/T \rfloor T)$$

This expression allows a computation of the transient distribution at any time $t \in \mathbb{R}^+$ without needing to integrate over ranges larger than the period T . For computing the remaining terms $P(0, s)$ with $s \leq T$, one can use the following iteration as given in Bellman [19], p.168: Starting with $I_0(u) := I$ for all $u \leq s$, the iteration

$$I_{n+1}(u) := \int_0^u I_n(v) Q(v) dv + I$$

leads to the limit

$$P(0, s) = \lim_{n \rightarrow \infty} I_n(s)$$

for all $s \leq T$.

Example A.4 Assume that the generator function is given by $Q(t) := (1 + \sin(\frac{2\pi}{T}t))Q$ for all $t \in \mathbb{R}_0^+$, with some constant generator Q . Then the terms $P(0, s)$ can be expressed by

$$\begin{aligned} P(0, s) &= \exp\left(\left(s + \int_0^s \sin\left(\frac{2\pi}{T}u\right) du\right) \cdot Q\right) \\ &= \exp\left(\left(s + \frac{T}{2\pi} \left(1 - \cos\left(\frac{2\pi}{T}s\right)\right)\right) \cdot Q\right) \end{aligned}$$

for all $s \leq T$, according to theorem A.21. Hence the transient distribution at time $t \in \mathbb{R}^+$ with initial distribution π is given by

$$X_t^\pi = \int d\pi (e^{T \cdot Q})^{\lfloor t/T \rfloor} \exp\left(\left(s + \frac{T}{2\pi} \left(1 - \cos\left(\frac{2\pi}{T}s\right)\right)\right) \cdot Q\right)$$

for $s := t - \lfloor t/T \rfloor \cdot T$.

The next theorem contains a recursion formula which yields a structural decomposition of the transition probabilities. Define

$$P_k(n) := P^{(k)}(0, nT) = \underbrace{\int_0^{nT} \int_0^{u_k} \dots \int_0^{u_2}}_{k \text{ integrals}} Q(u_1) \dots Q(u_k) du_1 \dots du_k$$

for all $k, n \in \mathbb{N}_0$. Now we can prove

Theorem A.26 *The following recursion formula holds for all $k, n \in \mathbb{N}_0$:*

$$P_k(n+1) = \sum_{j=0}^k P_{k-j}(n) P_j(1) \tag{A.5}$$

Proof: By definition, the partition

$$\begin{aligned} P_k(n+1) &= P_k(n) + \int_{nT}^{(n+1)T} \int_0^{u_k} \dots \int_0^{u_2} Q(u_1) \dots Q(u_k) du_1 \dots du_k \\ &= P_k(n) + \int_{nT}^{(n+1)T} P_{k-1}(n) Q(u_k) du_k \\ &\quad + \int_{nT}^{(n+1)T} \int_{nT}^{u_k} \dots \int_0^{u_2} Q(u_1) \dots Q(u_k) du_1 \dots du_k \\ &= \dots \\ &= \sum_{j=0}^k \underbrace{\int_{nT}^{(n+1)T} \int_{nT}^{u_j} \dots \int_{nT}^{u_2}}_{j \text{ integrals}} P_{k-j}(n) Q(u_1) \dots Q(u_j) du_1 \dots du_j \end{aligned}$$

holds for all $k, n \in \mathbb{N}_0$, defining the case of zero integrals as $P_k(n)$. Exploiting the periodicity of the generator $Q(t)$ leads to formula A.5.

☺

A. Markov Jump Processes

The initial values for this recursion are $P_0(n) = I$ for all $n \in \mathbb{N}_0$ and $P_k(0) = \delta_{k0} \cdot I$ for all $k \in \mathbb{N}_0$. This results in first iterates

$$P_1(n) = \int_0^{nT} Q(u) du$$

and

$$P_k(1) = \int_0^T \int_0^{u_k} \dots \int_0^{u_2} Q(u_1) \dots Q(u_k) du_1 \dots du_k$$

for all $k, n \in \mathbb{N}$.

Before examining the asymptotic behaviour of periodic Markov jump processes, one needs to define the term asymptotic distribution for periodic processes. The norm to be used will be the total variation as defined in definition A.10.

Definition A.13 Let X denote a periodic Markov jump process with period T . A family $(q_s : s \in [0, T[)$ of probability distributions shall be called a **periodic family of asymptotic distributions** if

$$\|X_{nT+s}^\pi - q_s\| \rightarrow 0 \text{ as } n \rightarrow \infty$$

for all $s \in [0, T[$, independently from the initial distribution π .

Theorem A.27 *If it exists, the periodic family $(q_s : s \in [0, T[)$ of asymptotic distributions is unique.*

Proof: Assume that $(q'_s : s \in [0, T[)$ is another periodic family of asymptotic distributions. Then for every $s \in [0, T[$ and an arbitrary probability measure π ,

$$\|q_s - q'_s\| \leq \|q_s - X_{nT+s}^\pi\| + \|q'_s - X_{nT+s}^\pi\| \rightarrow 0$$

as $n \rightarrow \infty$.

☺

Now the main result can be given. It reduces the existence of an asymptotic distribution for the periodic Markov jump process X to the existence of an asymptotic distribution for the embedded Markov chain Y at multiples of the period length. Define $Y = (Y_n : n \in \mathbb{N})$ as the homogeneous Markov chain with transition matrix $P(0, T)$ and let $Y^\pi = (Y_n^\pi : n \in \mathbb{N})$ denote the version of Y with initial distribution π .

Theorem A.28 *Let X denote a periodic Markov jump process with period T . If X has a periodic family of asymptotic distributions, then Y has a stationary distribution p . If Y has an asymptotic distribution q , then X has a periodic family of asymptotic distributions which is uniquely determined by*

$$q_s = \int dq P(0, s)$$

for all $s \in [0, T[$.

Proof: Let π denote the initial distribution of X at time $t_0 := 0$ and assume that $(q_s : s \in [0, T[)$ is a periodic family of asymptotic distributions. A necessary property of q_0 is

$$q_0 = \lim_{n \rightarrow \infty} \int d\pi P(0, nT) = \left(\lim_{n \rightarrow \infty} \int d\pi P(0, (n-1)T) \right) P(0, T) = \int dq_0 P(0, T)$$

which means that $p = q_0$ is a stationary distribution of Y .

Now let q be the asymptotic distribution of Y . Then

$$q_0 = \lim_{n \rightarrow \infty} \int d\pi P(0, nT) = \lim_{n \rightarrow \infty} \int d\pi P(0, T)^n = \lim_{n \rightarrow \infty} Y_n^\pi = q$$

does exist.

Using this distribution q , the periodic family of asymptotic distributions $(q_s : s \in [0, T[)$ is given by

$$q_s = \lim_{n \rightarrow \infty} \int d\pi P(0, nT) P(nT, nT + s) = \left(\lim_{n \rightarrow \infty} \int d\pi P(0, nT) \right) P(0, s) = \int dq P(0, s)$$

☺

Remark A.11 If Y can be shown to be ϕ -recurrent for some σ -finite measure ϕ on \mathcal{S} (see Orey [76], p.4), then the above theorem states: X has a periodic family of asymptotic distributions if and only if Y is positive recurrent with respect to ϕ (see Orey [76], p.30).

Example A.5 Resuming example A.4 and further assuming that Q is ϕ -recurrent and there is a measure q with $qQ = 0$ (which implies $qe^{T \cdot Q} = q$), we get a periodic family of asymptotic distributions $(q_s : s \in [0, T[)$ given by

$$q_s = \int dq \exp \left(\left(s + \frac{T}{2\pi} \left(1 - \cos \left(\frac{2\pi}{T} s \right) \right) \right) \cdot Q \right)$$

for all $s \in [0, T[$.

Concluding this section, the following theorems give sufficient conditions for aperiodicity and uniform recurrence (see Orey [76], pp.15,26) of the embedded Markov chain Y in terms of the jump kernels $(P(t) : t \in [0, T[)$ and the jump rates $(\gamma(t; x) : t \in [0, T[, x \in S)$.

Theorem A.29 Let X denote a periodic Markov jump process with period T and bounded jump rates $0 < m \leq \gamma(t; x) := -q(t; x, \{x\}) \leq M < \infty$ for all $x \in S$ and $t \in \mathbb{R}_0^+$. Be $(x, A) \in S \times \mathcal{S}$ with $\phi(A) > 0$. If there is a number $k \in \mathbb{N}_0$ with

$$\underbrace{\int_0^T \int_{u_1}^T \dots \int_{u_{k-1}}^T (P(u_1) \dots P(u_k))(x, A) du_1 \dots du_k}_{k \text{ integrals}} > 0 \quad (\text{A.6})$$

then

$$P(0, T)(x, A) > 0$$

A. Markov Jump Processes

Proof: Using formula A.1 for the transition probability $P(0, T)$, which has only positive components, we have

$$\begin{aligned}
P_b^{(k)}(0, T; x, A) &= \underbrace{\int_0^T \int_{u_1}^T \dots \int_{u_{k-1}}^T}_{k \text{ integrals}} \underbrace{\int_S \dots \int_S \int_A}_{k \text{ integrals}} e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x, dx_1) \dots \\
&\quad \dots e^{-\int_{u_{k-1}}^{u_k} \gamma(u; x_{k-1}) du} \gamma(u_k; x_{k-1}, dx_k) e^{-\int_{u_k}^t \gamma(u; x_k) du} du_k \dots du_1 \\
&= \underbrace{\int_0^T \int_{u_1}^T \dots \int_{u_{k-1}}^T}_{k \text{ integrals}} \underbrace{\int_S \dots \int_S \int_A}_{k \text{ integrals}} e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x) p(u_1; x, dx_1) \dots \\
&\quad \dots e^{-\int_{u_{k-1}}^{u_k} \gamma(u; x_{k-1}) du} \gamma(u_k; x_{k-1}) p(u_k; x_{k-1}, dx_k) e^{-\int_{u_k}^t \gamma(u; x_k) du} du_k \dots du_1
\end{aligned} \tag{A.7}$$

Remembering $0 < m \leq \gamma(t; x) \leq M < \infty$, the term

$$\begin{aligned}
&e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x) \dots e^{-\int_{u_{k-1}}^{u_k} \gamma(u; x_{k-1}) du} \gamma(u_k; x_{k-1}) e^{-\int_{u_k}^t \gamma(u; x_k) du} \\
&\geq e^{-M(t-s)} \gamma(u_1; x) \dots \gamma(u_k; x_{k-1})
\end{aligned}$$

is a continuous function of the vector $(\gamma(u_1; x), \dots, \gamma(u_k; x_{k-1}))$ with positive values. Hence, it assumes a positive minimum at some point in the compact set $[m, M]^k$. It follows that

$$P(0, T; x, A) \geq P_b^{(k)}(0, T; x, A) > 0$$

☺

Remark A.12 If for all $(x, A) \in S \times \mathcal{S}$ with $\phi(A) > 0$ there is a number $k \in \mathbb{N}_0$ such that condition A.6 holds, then it follows by the above theorem that Y is aperiodic and ϕ -irreducible. In this case, Y has a C-set according to Orey [76], pp.7–10.

Theorem A.30 Let X denote a periodic Markov jump process with bounded jump rates $0 < m \leq \gamma(t; x) \leq M < \infty$ for all $x \in S$ and $t \in \mathbb{R}_0^+$. If for any set $A \in \mathcal{S}$ with $\phi(A) > 0$ there is a number $K \in \mathbb{N}$ such that

$$\sum_{k=0}^K \underbrace{\int_0^T \int_{u_1}^T \dots \int_{u_{k-1}}^T}_{k \text{ integrals}} (P(u_1) \dots P(u_k))(x, A) du_1 \dots du_k > \varepsilon$$

for a positive $\varepsilon > 0$ and for all $x \in S$, then Y is aperiodic and uniformly ϕ -recurrent.

Proof: By theorem A.29 and remark A.12, Y is aperiodic. Choose any set $A \in \mathcal{S}$ with $\phi(A) > 0$. By assumption, there is a number $K \in \mathbb{N}$ such that

$$\sum_{k=0}^K \underbrace{\int_0^T \int_{u_1}^T \dots \int_{u_{k-1}}^T}_{k \text{ integrals}} \underbrace{\int_S \dots \int_S \int_A}_{k \text{ integrals}} p(u_1; x, dx_1) \dots p(u_k; x_{k-1}, dx_k) du_1 \dots du_k > \varepsilon$$

for a fixed $\varepsilon > 0$ and for all $x \in S$. For every $k \in \mathbb{N}_0$, the term

$$\begin{aligned} & e^{-\int_s^{u_1} \gamma(u; x) du} \gamma(u_1; x) \dots e^{-\int_{u_{k-1}}^{u_k} \gamma(u; x_{k-1}) du} \gamma(u_k; x_{k-1}) e^{-\int_{u_k}^t \gamma(u; x_k) du} \\ & \geq e^{-M(t-s)} \gamma(u_1; x) \dots \gamma(u_k; x_{k-1}) \end{aligned}$$

is a continuous function of the vector $(\gamma(u_1; x), \dots, \gamma(u_k; x_{k-1}))$ with positive values and assumes a minimum at some point in the compact set $[m, M]^k$. Denote this minimum by I_k and set $I := \min_{1 \leq k \leq K} I_k$. Note that $0 < I \leq 1$ and I is independent of $x \in S$. Then by equation A.7, it follows that

$$\begin{aligned} P(0, T; x, A) & \geq \\ & \geq \sum_{k=0}^K I_k \cdot \underbrace{\int_0^T \int_{u_1}^T \dots \int_{u_{k-1}}^T}_{k \text{ integrals}} \underbrace{\int_S \dots \int_S \int_A}_{k \text{ integrals}} p(u_1; x, dx_1) \dots p(u_k; x_{k-1}, dx_k) du_1 \dots du_k \\ & \geq I \cdot \sum_{k=0}^K \underbrace{\int_0^T \int_{u_1}^T \dots \int_{u_{k-1}}^T}_{k \text{ integrals}} (P(u_1) \dots P(u_k))(x, A) du_1 \dots du_k > I \cdot \varepsilon > 0 \end{aligned}$$

for all $x \in S$. Now the probability for Y of never reaching the set A from a state $x \in S$ is

$$\begin{aligned} \lim_{n \rightarrow \infty} P_x(Y_k \in A^c \forall k \in \{1, \dots, n\}) & = \lim_{n \rightarrow \infty} \underbrace{\int_{A^c} \dots \int_{A^c}}_{n \text{ integrals}} P(0, T; x, dx_1) \dots P(0, T; x_{n-1}, dx_n) \\ & \leq \lim_{n \rightarrow \infty} (1 - I \cdot \varepsilon)^n = 0 \end{aligned}$$

This means that starting from any state $x \in S$ the set A will be reached with probability 1. Furthermore, the probability

$$P_x\left(\bigcup_{k=1}^n \{Y_k \in A\}\right) \geq 1 - (1 - \varepsilon)^n$$

converges to 1 uniformly with respect to $x \in S$. This completes the proof.

☺

Remark A.13 Uniform ϕ -recurrence implies positive recurrence with respect to ϕ and a geometric convergence rate towards the asymptotic distribution of Y (see Orey [76], p.31).

A. *Markov Jump Processes*

B. Stochastic Point Fields

Stochastic point fields can be regarded as stochastic point processes with parameter sets not necessarily being the time axis (as the word "process" hints at) but a general space. They are a very useful concept in many application fields such as astronomy (Babu, Feigelson [8]), ecology (Thompson [87], Barnett, Turkman [11, 12]), the modelling of populations (Moyal [69]), rain fall (Cowpertwait [31], Phelan [79]) and mobile communication networks (Baum [16, 17]). An extensive list of references is given in Cressie [32]. Although in the rest of this thesis the focus shall rest on spatial queues arising from applicational demands in the field of mobile communication networks, it should not be forgotten that the use in other fields of application may turn out to be fruitful as well.

In spatial queueing theory, point fields can be used as a model for spatially distributed batch arrivals into a queue. Since arrivals in queues are always finite, we shall restrict ourselves to the theory of finite point fields. This section collects some classical results on stochastic point fields by Moyal [69] and Daley, Vere-Jones [33]. The first two subsections contain two alternative but equivalent ways of defining point fields. Subsection B.3 describes the construction of point fields from a given set of finite-dimensional marginal distributions. Finally, examples of point fields are given in the last subsection.

B.1. Set Version

A point field can be defined in two ways which are both useful for analyzing spatial arrival processes and spatial queues respectively. The most immediate way is that of defining a probability measure on the set of all finite point families.

Definition B.1 Let R denote an arbitrary space. A **point family** on R is a collection $\{x_1, \dots, x_n\}$ of elements $x_1, \dots, x_n \in R$ with the possibility of $x_i = x_j$ for $i \neq j \in \{1, \dots, n\}$.

Remark B.1 In difference to a set, a family may contain some elements more than once. In terms of points in a field, this means that there can be several points at the same position.

In order to define a measure on the set of all finite point families, a suitable σ -algebra needs to be constructed. This must provide for the possibility of several points having the same location as well as for the indistinguishability of points.

B. Stochastic Point Fields

Let $\mathcal{R} := \sigma(R)$ denote a σ -algebra on R and for $n \in \mathbb{N}$, let

$$\mathcal{R}^n := \bigotimes_{k=1}^n \mathcal{R}$$

denote the n -th product σ -algebra of \mathcal{R} on R^n . Further, let $R^0 := \{\emptyset\}$ and \mathcal{R}^0 denote the trivial σ -algebra on R^0 . Define

$$\mathcal{R}^\cup := \sigma \left(\bigcup_{n=0}^{\infty} \mathcal{R}^n \right)$$

as the minimal σ -algebra on $R^\cup := \bigcup_{n=0}^{\infty} R^n$ containing all sets in $\bigcup_{n=0}^{\infty} \mathcal{R}^n$.

Theorem B.1 *The σ -algebra \mathcal{R}^\cup consists of all subsets $A = \sum_{n=0}^{\infty} A_n$ of R^\cup such that $A_n \in \mathcal{R}^n$ for every $n \in \mathbb{N}_0$.*

Proof: see Moyal [69], p.3f

Since a family of points is seen not as an ordered n -tuple but its members are indistinguishable from one another, we need to restrict the σ -algebra \mathcal{R}^\cup on R^\cup to symmetric families only. Thus, by defining

$$\mathcal{R}_S^\cup := \mathcal{R}^\cup \cap \{A \in \mathcal{R}^\cup : A \text{ symmetric}\}$$

a suitable σ -algebra for a measure on finite point families has been found. Now, any probability measure Π_S on \mathcal{R}_S^\cup determines a probability space $(R^\cup, \mathcal{R}_S^\cup, \Pi_S)$ which is suitable for describing a point field.

B.2. Version for Counting Functions

Another way to see point families is in terms of counting functions. A counting function yields for every subset A of the basic space R the number of elements of the point family that are contained in A .

In this section, a point field shall be defined in terms of counting functions. After formally defining counting functions, a bijection between point families and counting functions shall be constructed. This bijection will be used in order to transfer the point field definition from the setting of point families into terms of counting functions.

Definition B.2 Let $(R, \sigma(R))$ be a measurable space with $\{x\} \in \sigma(R)$ for every $x \in R$. A function $N : \sigma(R) \rightarrow \mathbb{N}_0$ is called **counting function** if for any collection $(A_i : i \in I)$ of disjoint measurable subsets of R ,

1. there are at most finitely many subsets A_{i_1}, \dots, A_{i_n} for which $N(A_{i_k}) \geq 1$.

2. the equality

$$N \left(\sum_{i \in I} A_i \right) = \sum_{i \in I} N(A_i) = \sum_{k=1}^n N(A_{i_k})$$

holds.

Let \mathcal{C} denote the space of all counting functions on R .

Remark B.2 If the number of points in the measurable set $A \in \sigma(R)$ is represented by $N(A)$, then any family of points has a unique representation by a counting function N . This is only true because of the restriction that every singleton $\{x\}$ is contained in the σ -algebra $\sigma(R)$ (see Moyal [69], p.9).

Now the mentioned bijection between the space of finite point families and the class of counting functions shall be constructed. Define the **Dirac measure** at $x \in R$ by

$$\delta_x(A) := \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for } x \notin A \end{cases}$$

for all $A \in \sigma(R)$.

Theorem B.2 For all $n \in \mathbb{N}$ and all $x_1, \dots, x_n \in R$, define

$$T(\{x_1, \dots, x_n\}) := \sum_{i=1}^n \delta_{x_i}$$

Then the mapping T is a bijection between the space of all finite point families on R and the class \mathcal{C} of counting functions on R .

Proof: see Moyal [69], p.7

Using this bijection T , a σ -algebra on \mathcal{C} can be defined by $\sigma(\mathcal{C}) := T(\mathcal{R}_S^\cup)$, i.e. by all sets of counting functions whose inverse image under T^{-1} belongs to \mathcal{R}_S^\cup .

Theorem B.3 The σ -algebra $\sigma(\mathcal{C}) = T(\mathcal{R}_S^\cup)$ is the smallest σ -algebra containing all sets of the form

$$\{N \in \mathcal{C} : N(A_i) = k_i \quad \forall i \in \{1, \dots, n\}\}$$

with $n \in \mathbb{N}$ and $A_i \in \sigma(R)$ measurable, $k_i \in \mathbb{N}_0$ for all $i \in \{1, \dots, n\}$.

Proof: see Moyal [69], p.8

Defining a probability measure $\Pi_{\mathcal{C}}$ on $\sigma(\mathcal{C})$ by $\Pi_{\mathcal{C}}(A) := \Pi_S(T^{-1}(A))$, the probability space $(\mathcal{C}, \sigma(\mathcal{C}), \Pi_{\mathcal{C}})$ is equivalent to $(R^\cup, \mathcal{R}_S^\cup, \Pi_S)$ and hence suitable to define a point field.

B. Stochastic Point Fields

Definition B.3 Any one of the equivalent probability spaces $(\mathcal{C}, \sigma(\mathcal{C}), \Pi_C)$ and $(R^\cup, \mathcal{R}_S^\cup, \Pi_S)$ shall be called **point field**. The first one will be referred to as the **version for counting functions**, the second one as **set version**.

Remark B.3 In the following, the standard version referred to shall be the version for counting functions, since it is easier to describe the main properties of point fields in terms of this version. Thus, if nothing else is stated, a point field shall be defined in terms of counting functions from now on.

B.3. Construction

In order to construct a point field on a space R from distributions on finitely many subsets of R , we will use results from Daley, Vere-Jones [33], chapter 6. In this, it is assumed that the space R be Polish (see Daley, Vere-Jones [33], p.153). Hence, the same assumptions are made here, although they might not be necessary.

Definition B.4 A complete separable metric space R is called a **Polish space**. The smallest σ -algebra containing all open subsets of R is called the **Borel σ -algebra** on R . A set which is an element of the Borel σ -algebra on R is called a **Borel set** of R .

Remark B.4 If R is a Polish space and $\sigma(R)$ the Borel σ -algebra on R , then every singleton $\{x\}$ is contained in $\sigma(R)$. This follows from

$$\{x\} = \bigcap_{n=1}^{\infty} B\left(x, \frac{1}{n}\right)$$

with $B(x, d) := \{y \in R : d(x, y) < d\}$ denoting the set of points $y \in R$ having a distance less than d from x . That means that the condition on $\sigma(R)$ appearing in definition B.2 is satisfied.

The first immediate result confirms the unique representation of a point field by a sufficiently large set of finite-dimensional marginal distributions.

Theorem B.4 Let R be a Polish space and $\sigma(R)$ the Borel σ -algebra on R . Define

$$\Pi(S_1, \dots, S_n)(k_1, \dots, k_n) := \Pi_C(\{N \in \mathcal{C} : N(S_i) = k_i \ \forall i \in \{1, \dots, n\}\})$$

Then the probability measure Π_C is uniquely determined by the marginal distributions $\Pi(S_1, \dots, S_n)$ on finite families of disjoint sets $S_1, \dots, S_n \in \sigma(R)$ from a semi-ring of bounded sets generating $\sigma(R)$.

Proof: see Daley, Vere-Jones [33], p.167

The main result states sufficient and necessary conditions for the existence of a point field, if the set of finite-dimensional marginal distributions is given. The first condition is merely

notational. The second condition is the consistency condition for Kolmogorov's extension theorem, which ensures that a probability measure having the given marginal distributions can be constructed. The last two conditions guarantee that the support set of the constructed probability measure is exactly the space \mathcal{C} of counting functions.

Theorem B.5 *Let R be a Polish space and $\sigma(R)$ the Borel σ -algebra on R . Further, let*

$$(\Pi(S_1, \dots, S_n) : n \in \mathbb{N}, S_1, \dots, S_n \in \sigma(R))$$

denote the set of all finite-dimensional distributions. Then there is a point field Π_C on $(R, \sigma(R))$ such that

$$\Pi(S_1, \dots, S_n)(k_1, \dots, k_n) = \Pi_C(\{N \in \mathcal{C} : N(S_i) = k_i \ \forall i \in \{1, \dots, n\}\}).$$

if and only if the following conditions are satisfied:

1. *For all $n \in \mathbb{N}$ and every permutation (i_1, \dots, i_n) of the integers $(1, \dots, n)$, the distributions*

$$\Pi(S_1, \dots, S_n) = \Pi(S_{i_1}, \dots, S_{i_n})$$

are equal.

2. *For every $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$, the consistency condition*

$$\Pi(S_1, \dots, S_{n-1})(k_1, \dots, k_{n-1}) = \sum_{k=0}^{\infty} \Pi(S_1, \dots, S_n)(k_1, \dots, k_{n-1}, k)$$

holds.

3. *For every $n \in \mathbb{N}$ and $S_1, \dots, S_n \in \sigma(R)$, the measure $\Pi(S_1, \dots, S_n)$ has support \mathbb{N}_0^n .*

4. *For disjoint Borel sets $A, B \in \sigma(R)$ and integers $k, l \in \mathbb{N}_0$, the equality*

$$\Pi(A, B, A + B)(k, l, k + l) = \Pi(A, B)(k, l)$$

holds.

Proof: The necessity is shown first: Condition 1 is a mere notational requirement. Condition 2 expresses the consistency requirement needed in Kolmogorov's extension theorem (see Gikhman, Skorokhod [44], p.108 (theorem 3)). If it were not satisfied, the given distributions could not be marginal distributions of the same measure. Condition 3 is necessary, because if for some measure $\Pi(S_1, \dots, S_n)$ the support would exceed \mathbb{N}_0^n , then the support of any measure on $(R, \sigma(R))$ with marginal measure $\Pi(S_1, \dots, S_n)$ would exceed $\mathbb{N}_0^{\sigma(R)}$ and hence the space $\mathcal{C} \subset \mathbb{N}_0^{\sigma(R)}$ of counting functions. Finally, if condition 4 fails to hold, any measure on $(R, \sigma(R))$ with marginal measures $\Pi(A, B, A + B)$ and $\Pi(A, B)$ would have a support set

B. Stochastic Point Fields

that violates the second property in definition B.2 and hence is greater than the set of counting functions.

Now we show sufficiency: According to Kolmogorov's extension theorem (see Gikhman, Skorokhod [44], p.108 (theorem 3)), the first two conditions guarantee that a probability measure Π_C on $(R, \sigma(R))$ with marginal measure $\Pi(S_1, \dots, S_n)$ can be constructed. Condition 3 implies that the support of Π_C is on $\mathcal{N}_0^{\sigma(R)}$ and further

$$\sum_{n=0}^{\infty} \Pi(R)(n) = 1$$

which ensures the first property in definition B.2. Condition 4 states that for any two disjoint Borel sets $A, B \in \sigma(R)$,

$$\Pi_C(\{N \in \mathcal{C} : N(A) + N(B) = N(A + B)\}) = 1$$

Since the Borel σ -algebra of a Polish space can be generated by a countable semi-ring (see Daley, Vere-Jones [33], p.608 (theorem A2.1.III)), this implies that the second property in definition B.2 holds almost surely with respect to Π_C . Hence, Π_C is a probability measure on the set \mathcal{C} of counting functions.

☺

B.4. Examples

Concluding this section, some examples of stochastic point fields are given. The first one is the well-known Poisson field, which assumes independence of the point positions. The next two examples are widely used variations of the Poisson field. Example B.4 is rather a general framework for the design of finite point fields than a specific example. For more examples see Cressie [32], Upton, Fingleton [92] or Diggle [35].

Example B.1 Let $(R, \sigma(R))$ be a measurable space with $\{x\} \in \sigma(R)$ for every $x \in R$. If the distribution of points in disjoint sets is independent, i.e. if

$$\Pi(S_1, \dots, S_n)(k_1, \dots, k_n) = \prod_{i=1}^n \Pi(S_i)(k_i)$$

for any disjoint sets $S_1, \dots, S_n \in \sigma(R)$, and if for all sets $S \in \sigma(R)$ the distribution $\Pi(S)$ is a Poisson distribution with some parameter $\mu(S) \in [0, \infty]$, then the point field $(\mathcal{C}, \sigma(\mathcal{C}), \Pi_C)$ shall be called a **Poisson point field** on $(R, \sigma(R))$. These are examined in Kingman [55]. Some important basic properties are given in the following theorems.

Theorem B.6 *The function $\mu : \sigma(R) \rightarrow [0, \infty]$, which yields the parameters $\mu(S)$ of the Poisson distributions $\Pi(S)$, is a measure on $(R, \sigma(R))$. On the other hand, if μ is a non-atomic measure on $(R, \sigma(R))$ which can be expressed in the form*

$$\mu = \sum_{n=1}^{\infty} \mu_n$$

with $\mu_n(R) < \infty$ for every $n \in \mathbb{N}$, then there is a Poisson point field $(\mathcal{C}, \sigma(\mathcal{C}), \Pi_{\mathcal{C}})$ on $(R, \sigma(R))$ with $\Pi(S)$ having parameter $\mu(S)$.

Proof: see Kingman [55], p.11f,23

Since $E(\Pi(S)) = \mu(S)$, the measure μ is called the **mean measure** of $(\mathcal{C}, \sigma(\mathcal{C}), \Pi_{\mathcal{C}})$.

Theorem B.7 Let μ be a non-atomic finite measure on $(R, \sigma(R))$. Further, let $f : R \rightarrow \mathbb{R}_0^+$ be a bounded μ -measurable function. Then the measure ν defined by

$$\nu(A) = \int_A f(x) d\mu(x)$$

for all $A \in \sigma(R)$ is a mean measure in the sense of theorem B.6.

Proof: It is immediate to check the assumptions of theorem B.6.

The function f shall be called the **intensity function** with respect to μ . The value $f(x)$ is called **intensity** of $x \in R$.

Example B.2 If the mean measure of a Poisson point field is allowed to be random in the same probability space as the point field itself, then the resulting point field shall be called a **Cox field**. For a more formal definition and properties of Cox fields, see Cressie [32], pp.657-661, or Grandell [45].

Example B.3 Another point field which is derived from the Poisson point field is the **Poisson cluster field**. It is defined by the following construction:

1. Parent events are realized from an inhomogeneous Poisson point field on R with some mean measure μ .
2. Each parent produces a random number of offspring.
3. The offspring are positioned arbitrarily in R .
4. The final process is composed of the superposition of offspring only.

A special kind of Poisson cluster fields is the **Neyman–Scott process**. A Poisson cluster field on $R = \mathbb{R}^d$ is called a Neyman–Scott process if

1. The random number K of offspring is realized independently and identically for each parent according to a discrete probability distribution $(p_k : k \in \mathbb{N}_0)$.
2. The positions of the offspring relative to their parents are independently and identically distributed according to a d -dimensional density function $f : \mathbb{R}^d \rightarrow \mathbb{R}_0^+$.

For properties of Neyman–Scott processes, see Cressie [32], pp.662-669.

B. Stochastic Point Fields

Example B.4 An easy and constructive way to specify finite point fields can be found in Daley, Vere-Jones [33], chapter 5.3. By this method, first the probability p_n of the number n of points in R is given and then, conditioned on $n \geq 1$, a probability distribution Π_n of the locations of the points. The indistinguishability implies that every Π_n must be symmetric.

Theorem B.8 *Let $(R, \sigma(R))$ be a Polish space with its Borel σ -algebra. Then a finite point field is given by a discrete probability distribution $(p_n : n \in \mathbb{N}_0)$ and, for each integer $n \in \mathbb{N}$, a symmetric probability distribution Π_n on the σ -algebra $\sigma(R)^n$ of R^n .*

Proof: see Daley, Vere-Jones [33], p.126 (proposition 5.3.II)

Bibliography

- [1] A. Albert (1961): "Estimating the Infinitesimal Generator of a Continuous Time, Finite State Markov Process", *Ann. Math. Statist.* 38, pp.727-753
- [2] T.W. Anderson, L.A. Goodman (1957): "Statistical inference about Markov chains", *Ann. Math. Statistics* 28, pp.89-110
- [3] E. Arjas, T.P. Speed (1973): "Topics in Markov-Additive Processes", *Math. Scandinav.* 33, pp.171-192
- [4] S. Asmussen (1987): "Applied Probability and Queues", John Wiley & Sons (Chichester)
- [5] S. Asmussen, H. Thorisson (1987): "A Markov Chain Approach to Periodic Queues", *J. Appl. Prob.* 24, pp.215-225
- [6] S. Asmussen, M. Olsson, O. Nerman (1996): "Fitting phase-type distributions via the EM algorithm", *Scand. J. Statist.* 23, No.4, pp.419-441
- [7] S. Asmussen (1997): "Phase-Type Distributions and Related Point Processes: Fitting and Recent Advances", in: A. Alfa, S. Chakravarthy: "Matrix-Analytical Methods in Stochastic Models", Marcel Dekker, pp.137-149
- [8] G. Babu, E. Feigelson (1996): "Spatial Point Processes in Astronomy", *Journal of Statistical Planning and Inference* 50, pp.311-326
- [9] F. Baccelli, S. Zuyev (1997): "Stochastic Geometry Models of Mobile Communication Networks", in: J. Dshalalow: "Frontiers in Queueing", CRC Press, pp.227-243
- [10] N. Bambos, J. Walrand (1989): "On Queues with periodic inputs", *J. Appl. Prob.* 26, pp.381-389
- [11] V. Barnett, K.F. Turkman (1993): "Statistics for the Environment", John Wiley & Sons
- [12] V. Barnett, K.F. Turkman (1994): "Statistics for the Environment 2: Water Related Issues", John Wiley & Sons
- [13] H. Bauer (1991): "Wahrscheinlichkeitstheorie", Walter de Gruyter (Berlin)

Bibliography

- [14] D. Baum (1996): "Ein Faltungskalkül für Matrizenfolgen und verallgemeinerte Poisson-Gruppenprozesse", Research Report No.96-36, Department of Mathematics and Computer Science, University of Trier
- [15] D. Baum (1996): "Convolution Algorithms for BMAP/G/1 - Queues", Research Report No.96-22, Department of Mathematics and Computer Science, University of Trier
- [16] D. Baum (1998): "On Markovian Spatial Arrival Processes for the Performance Analysis of Mobile Communication Networks", Research Report No.98-07, Department of Mathematics and Computer Science, University of Trier
- [17] D. Baum (1998): "The Infinite Server Queue With Markov Additive Arrivals in Space", Research Report No.98-31, Department of Mathematics and Computer Science, University of Trier
- [18] A. Beck (1963): "On the strong law of large numbers", in: F. Wright: "Ergodic Theory", Academic Press (New York), pp.21-53
- [19] R. Bellman (1997): "Introduction to Matrix Analysis", SIAM
- [20] R. Botta, C. Harris, W. Marchal (1987): "Characterizations of Generalized Hyperexponential Distribution Functions", Commun. Statist. - Stochastic Models 3(1), pp.115-148
- [21] R. Boucherie, N. Dijk: "On a queueing network model for cellular mobile telecommunication networks", manuscript
- [22] R. Boucherie, M. Mandjes: "Estimation of performance measures for product form cellular mobile communications networks", manuscript
- [23] N. Bourbaki (1969): "Intégration", Diffusion C.C.L.S. (Paris)
- [24] D. Chauveau, C. Martin, A. Rooij, F. Ruymgaart (1996): "Discrete Signed Mixtures of Exponentials", Commun. Statist. - Stochastic Models 12(2), pp.245-263
- [25] E. Çinlar (1972): "Markov Additive Processes I,II", Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete 24, pp.85-93,95-121
- [26] E. Çinlar (1975): "Introduction to Stochastic Processes", Prentice-Hall (Englewood Cliffs, NJ)
- [27] E. Çinlar (1975): "Levy Systems of Markov Additive Processes", Z. Wahrscheinlichkeitstheorie verw. Geb. 31, pp.175-185
- [28] E. Çinlar (1977): "Markov Additive Processes and Semi-Regeneration", Proc. 5th Conf. Probab. Theory, Brasov 1974, pp.33-49

- [29] E. Çinlar (1982): "Regenerative systems and Markov additive processes", Seminar on Stochastic Processes in Evanston, Illinois, March 1982, Prog. Probab. Stat. 5, pp.123-147
- [30] E. Çinlar (1995): "An introduction to spatial queues", in: Dshalalow: "Advances in Queueing", CRC Press, pp.103-118
- [31] P. Cowpertwait (1995): "A generalized spatial-temporal model of rainfall based on a clustered point process", Proc. R. Soc. Lond., Ser. A 450, No.1938, pp.163-175
- [32] N. Cressie (1993): "Statistics for Spatial Data", John Wiley & Sons
- [33] D. Daley, D. Vere-Jones (1988): "An Introduction to the Theory of Point Processes", Springer
- [34] A.P. Dempster, N.M. Laird, D.B. Rubin (1977): "Maximum Likelihood from Incomplete Data via the EM Algorithm", J. R. Stat. Soc., Ser. B39, pp.1-38
- [35] Diggle, Peter J. (1983): "Statistical analysis of spatial point patterns", Academic Press
- [36] J. Doob (1953): "Stochastic Processes", John Wiley & Sons
- [37] I.I. Ezhov, A.V. Skorokhod (1969): "Markov processes with homogeneous second component", Theory of Probability and its Applications 14, pp.1-13, 652-667
- [38] G.I. Falin (1989): "Periodic Queues in Heavy Traffic", Adv. Appl. Prob. 21, pp.485-487
- [39] G. Fayolle, V. Malyshev, M. Menshikov (1995): "Topics in the Constructive Theory of Countable Markov Chains", Cambridge University Press
- [40] F.G. Foster (1951): Discussion to Kendall [54]
- [41] F.G. Foster (1953): "On Stochastic Matrices Associated with Certain Queueing Processes", Ann. Math. Stat. 26, pp.355-360
- [42] J. Gibson (1996): "The Mobile Communication Handbook", CRC Press (Boca Raton)
- [43] M. Gilbert (2000): "Zur Parameterschätzung für eine Klasse von Markov-additiven Prozessen", diploma thesis, University of Trier
- [44] I. Gikhman, A. Skorokhod (1969): "Introduction to the Theory of Random Processes", W.B. Saunders Company
- [45] J. Grandell (1976): "Doubly Stochastic Poisson Process", Lecture Notes in Mathematics, No.529, Springer (Berlin)

Bibliography

- [46] J.M. Harrison, A.J. Lemoine (1977): "Limit Theorems for Periodic Queues", *J. Appl. Prob.* 14, pp.566-576
- [47] A.M. Hasofer (1964): "On the Single-Server Queue with Non-Homogeneous Poisson Input and General Service Time", *J. Appl. Prob.* 1, pp.369-384
- [48] H. Herrlich (1986): "Topologie I: Topologische Räume", Heldermann Verlag (Berlin)
- [49] D.P. Heyman, W. Whitt (1984): "The Asymptotic Behavior of Queues with Time-Varying Arrival Rates", *J. Appl. Prob.* 21, pp.143-156
- [50] J. Hofmann (1998): "The BMAP/G/1 queue with level dependent arrivals", dissertation at the University of Trier, Dept. of Mathematics and Computer Science
- [51] M. Jamshidian, R.I. Jennrich (1997): "Acceleration of the EM Algorithm by using Quasi-Newton Methods", *J. R. Statist. Soc. B*59, pp.569-587
- [52] N. Jewell (1982): "Mixtures of Exponential Distributions", *The Annals of Statistics* 10(2), pp.479-484
- [53] E. Kamke (1977): "Differentialgleichungen", vol.1, Teubner
- [54] D. Kendall (1951): "Some problems in the theory of queues", *J. Roy. Statist. Soc. B*13, pp.151-185
- [55] J. Kingman (1993): "Poisson Processes", Clarendon Press
- [56] A.J. Lemoine (1981): "On Queues with Periodic Poisson Input", *J. Appl. Prob.* 18, pp.889-900
- [57] A.J. Lemoine (1989): "Waiting Time and Workload in Queues with Periodic Poisson Input", *J. Appl. Prob.* 26, pp.390-397
- [58] B.G. Leroux (1992): "Maximum-Likelihood Estimation for Hidden Markov Chains", *Stoch. Proc. Appl.* 40, pp.127-143
- [59] R.Sh. Liptser, A.N. Shirayev (1989): "Theory of Martingales", Kluwer
- [60] D. Lucantoni (1991): "New Results on the Single Server Queue with a Batch Markovian Arrival Process", *Comm. Statist. - Stochastic Models* 7(1), pp.1-46
- [61] P. Marlin (1973): "On the ergodic theory of Markov chains", *Operat. Res.* 21, pp.617-622
- [62] J. Mauldon (1958): "On non-dissipative Markov chains", *Proc. Camb. Phil. Soc.* 53, pp.825-835
- [63] W.A. Massey, W. Whitt (1993): "Networks of infinite server queues with non-stationary Poisson input", *Queueing Systems* 13, pp.183-250

- [64] G.J. McLachlan, T. Krishnan (1997): "The EM Algorithm and Extensions", John Wiley and Sons
- [65] B. Melamed, D. Yao (1995): "The ASTA property", in: J. Dshalalow: "Advances in Queueing", CRC Press (Boca Raton), pp.195-224
- [66] X.-L. Meng, D. Dyk (1997): "The EM Algorithm - an Old Folk-Song Sung to a Fast New Tune", J. R. Statist. Soc. B59, pp.511-567
- [67] C. Moler, C. van Loan (1978): "Nineteen dubious ways to compute the exponential of a matrix", SIAM Review 20, pp.801-836
- [68] E. Mourier (1953): "Elements Aleatoires dans un espace de Banach", Ann. Inst. H. Poincare 13, pp.161-244
- [69] J. Moyal (1962): "The General Theory of Stochastic Population Processes", Acta mathematica 108, pp.1-31
- [70] W. Nef (1977): "Lehrbuch der linearen Algebra", Birkhäuser (Basel)
- [71] M. Neuts (1979): "A versatile Markovian point process", J. Appl. Prob. 16, pp.764-79
- [72] M. Neuts (1989): "Structured Stochastic Matrices of M/G/1 Type and Their Applications", Marcel Dekker
- [73] J. Neveu (1961): "Une Generalisation des Processus a Croissements Positifs Independants", Abh. math. Sem. Univ. Hamburg 25, pp.36-61
- [74] J. Neveu (1969): "Mathematische Grundlagen der Wahrscheinlichkeitstheorie", Oldenbourg Verlag (München)
- [75] P. Ney, E. Nummelin (1987): "Markov additive processes I,II", Ann. Probab. 15, pp.561-592,593-609
- [76] S. Orey (1971): "Limit Theorems for Markov Chain Transition Probabilities", Van Nostrand (London) 1971
- [77] A. Pacheco, N.U. Prabhu (1995): "Markov-additive processes of arrivals", in: Dshalalow: "Advances in Queueing", CRC Press, pp.167-194
- [78] A.G. Pakes (1969): "Some Conditions for Ergodicity and Recurrence of Markov Chains", Oper. Res. 17, pp.1058-1061
- [79] M. Phelan (1991): "Point processes and inference for rainfall fields", Applied probability, Proc. Symp. Sheffield/UK 1990, IMS Lect. Notes, Monogr. Ser. 18, pp.127-148

Bibliography

- [80] V. Ramaswami (1988): "A Stable Recursion for the Steady State Vector in Markov Chains of M/G/1 Type", *Comm. Statist. - Stochastic Models* 4(1), pp.183–188
- [81] T. Rappaport (1996): "Wireless Communications", Prentice Hall
- [82] T. Rolski (1987): "Approximation of Periodic Queues", *Adv. Appl. Prob.* 19, pp.691-707
- [83] T. Rolski (1989): "Relationships between Characteristics in Periodic Poisson Queues", *Queueing Systems* 4, pp.17-26
- [84] Z. Rosberg (1981): "A note on the ergodicity of Markov chains", *J. Appl. Prob.* 18, pp.112–121
- [85] T. Ryden (1996): "An EM algorithm for estimation in Markov-modulated Poisson processes", *Comput. Stat. Data Anal.* 21, Nr.4, pp.431-447
- [86] B. Sevast'yanov (1957): "An Ergodic Theorem for Markov Processes and its Application to Telephone Systems with Refusals", *Theory of Probability and its Applications* 2, pp.104–112
- [87] H. Thompson (1955): "Spatial point processes, with applications to ecology", *Biometrika* 42, pp.102-115
- [88] D.M. Titterington (1984): "Recursive Parameter Estimation Using Incomplete Data", *J. R. Statist. Soc. B*46, pp.257-267
- [89] D.M. Titterington, A.F.M. Smith, U.E. Makov (1985): "Statistical Analysis of Finite Mixture Distributions", John Wiley and Sons
- [90] R. Tweedie (1975): "A relation between positive recurrence and mean drift for Markov chains", *Austral. J. Statist.* 17, pp.96–102
- [91] R. Tweedie (1976): "Criteria for classifying general Markov chains", *Adv. Appl. Prob.* 8, pp.737–771
- [92] G. Upton, B. Fingleton (1985): "Spatial data analysis by example. Volume 1: Point pattern and quantitative data", John Wiley & Sons
- [93] A. Viterbi (1995): "Principles of Spread Spectrum Communication", Addison-Wesley (Reading)
- [94] H. Willie (1998): "Periodic Steady State of Loss Systems with Periodic Inputs", *Adv. Appl. Prob.* 30, pp.152-166
- [95] R. Wolff (1989): "Stochastic Modeling and the Theory of Queues", Prentice Hall
- [96] S. Yang (1998): "CDMA RF System Engineering", Artech House (Boston)