Probability on Graphs and Semigroups: Abstract Reliability Theory

Kyle Siegrist
Department of Mathematical Sciences
University of Alabama in Huntsville
siegrist@math.uah.edu

Draft: October 20, 2020
© 2020 Kyle Siegrist
Permission is granted to copy and distribute this work under the terms of the GNU Public License.

To Delia and Heidi
Preface

The purpose of this text is to study abstract settings that generalize many of the basic concepts in classical reliability theory. So let us start with a brief review of the classical theory. For notation, $\mathbb{R}$ is the set of real numbers, $\mathbb{N}$ the set of nonnegative integers, and $\mathbb{N}_+$ the set of positive integers.

Standard Concepts

The set $S = [0, \infty)$ or the set $S = \mathbb{N}$, along with the ordinary order $\leq$, and the addition operation $+$ form the standard models of continuous time and discrete time, respectively. The addition operator allows us to shift forward in time, and the order relation allows for comparisons, both crucial for various aging concepts. Of course, $+$ and $\leq$ are intimately related since for $x, y \in S$, $x \leq y$ if and only if $x + t = y$ for some $t \in S$.

Another essential component is measure: Lebesgue measure $\lambda$ on the collection $\mathcal{S}$ of measurable subsets of $[0, \infty)$ or counting measure $\lambda$ on the collection $\mathcal{S}$ of all subsets of $\mathbb{N}$. Crucially, both are translation invariant

$$\lambda(x + A) = \lambda(A), \quad x \in S, A \in \mathcal{S}$$

providing another critical connection between the various structures.

Suppose now that $X$ is a random variable with values in $S$, representing the lifetime of a device (or an organism), and that $X$ has a density function $f$ (with respect to $\lambda$). The reliability function $F$ of $X$ is the right distribution function, defined by

$$F(x) = \mathbb{P}(X \geq x), \quad x \in S$$

so that $F(x)$ is the probability that the lifetime is at least $x$. The failure right rate function $r$ is defined by $r(x) = f(x)/F(x)$ for $x \in S$, so that roughly $r(x)$ is the rate (conditional density) of failure, given a lifetime of at least $x$. The device might have increasing, decreasing, or constant failure rate, depending on whether $r$ is increasing, decreasing, or constant on $S$. Or course, $r$ might have a more complicated behavior, for example decreasing initially, then constant for a while, and then increasing. Note that only the order relation is necessary for these concepts.

Other aging properties arise by comparing the conditional distribution of $X - x$ given $X \geq x$ with the distribution of $X$ for $x \in S$. That is, how does the distribution of the remaining life, given survival up to time $x$ compare with the original life distribution of $X$? Since the reliability function determines the distribution, one simple approach is to compare $\mathbb{P}(X \geq x + y \mid X \geq x)$, with $\mathbb{P}(X \geq y)$ for $x, y \in S$. The first is the probability that the device will last at least another $y$ units given survival up to time $x$ while the second is simply the probability of survival up to time $y$. From the simple definition of conditional probability, this is equivalent to comparing $F(x + y)$ with $F(x)F(y)$ for $x, y \in S$. If $F(x + y) = F(x)F(y)$ for all $x, y \in S$ then $X$ is memoryless. If $F(x + y) \leq F(x)F(y)$ for all $x, y \in S$, then $X$ is new better than used, and if $F(x + y) \geq F(x)F(y)$ for all $x, y \in S$ then $X$ is new worse than used. Again, of course, none of these might apply, so that the distribution of $X$ is more complicated. Note that the addition operation, as well as the associated order relation, are necessary for these concepts.

Suppose now that $U = (U_1, U_2, \ldots)$ is a sequence of independent, identically distributed random variables in $S$ with common density function $f$. There are two simple ways to construct a new sequence of variables. First let $X = (X_1, X_2, \ldots)$ be the partial sum sequence associated with $U$ so that $X_n = \sum_{i=1}^{n} U_i$ for $n \in \mathbb{N}_+$. If we think of $U$ as the sequence of interarrival times in a renewal process then $X$ is the corresponding sequence of arrival times. Another method is to construct the sequence of record variables $X = (X_1, X_2, \ldots)$ associated with $U$. Let $X_1 = U_1$. Next let $X_2 = U_{N_2}$ where $N_2 = \min\{n > 1 : U_n \geq X_1\}$ and continue in this way. Note that the addition operator $+$ is crucial in the first construction while the order relation $\leq$ is crucial
in the second. In both cases, if we think of $X$ as a random walk in the space $(S, \leq)$, then the corresponding point process $N = \{N_A : \text{measurable } A \subseteq S\}$ is defined by $N_A = \sum_{n=1}^{\infty} 1(X_n \in A)$, the number of points in $A$.

### Continuous Time

In the case of continuous time, we can give some specific and well-known results. Again, let $X$ be a random variable with values in $S = [0, \infty)$ and with density function $f$ (with respect to Lebesgue measure $\lambda$) and reliability function $F$. The expected value of $X$ can be obtained by integrating the reliability function:

$$
\int_0^\infty F(x) \, dx = E(X)
$$

A bit more generally, and not as well known,

$$
\int_0^\infty \frac{x^n}{n!} F(x) \, dx = E \left[ \frac{X^{n+1}}{(n+1)!} \right], \quad n \in \mathbb{N}
$$

(1)

Random variable $X$ has constant failure rate if and only if it is memoryless if and only if the conditional distribution of $X$ given $X \geq x$ is the same as the distribution of $X$ if and only if the distribution is exponential:

$$
f(x) = \alpha e^{-\alpha x}, \quad F(x) = e^{-\alpha x}, \quad x \in [0, \infty)
$$

where $\alpha \in (0, \infty)$ is the failure rate constant. When $X$ has this exponential distribution, the partial sum process and the record process corresponding to the distribution of $X$ are the same, and the sequence $X = (X_1, X_2, \ldots)$ gives the arrival times of the Poisson process with rate $\alpha$. In particular for $n \in \mathbb{N}_+$, $X_n$ has the gamma distribution with parameters $\alpha$ and $n$, with density function $f_n$ defined by

$$
f_n(x) = \alpha^n \frac{x^{n-1}}{(n-1)!} F(x), \quad x \in [0, \infty)
$$

(2)

Moreover, the Poisson process is the “most random” way to put points in $[0, \infty)$ in the sense that given $X_{n+1} = x \in [0, \infty)$, the random vector $(X_1, X_2, \ldots, X_n)$ is uniformly distributed on

$$
\{(x_1, x_2, \ldots, x_n) \in [0, \infty)^n : x_1 \leq x_2 \leq \cdots x_n \leq x\}
$$

a set that has measure $x^n/n!$. Note the repeated appearance of the expression $x^k/k!$ for various $x \in [0, \infty)$ and $k \in \mathbb{N}$.

### Discrete Time

In the case of discrete time, we can also give some specific and well-known results. Again, let $X$ be a random variable with values in $S = \mathbb{N}$ and with density function $f$ (with respect to counting measure $\#$) and reliability function $F$. Summing the reliability function gives a standard moment result:

$$
\sum_{x=0}^{\infty} F(x) = E(X) + 1
$$

A bit more generally, and not as well known,

$$
\sum_{x=0}^{\infty} \binom{x+n}{n} F(x) = E \left[ \frac{X+n+1}{n+1} \right], \quad n \in \mathbb{N}
$$

(3)

Random variable $X$ has constant failure rate if and only if it is memoryless if and only if conditional distribution of $X - x$ given $X \geq x$ is the same as the distribution of $X$ if and only if the distribution is geometric:

$$
f(x) = \alpha(1-\alpha)^x, \quad F(x) = (1-\alpha)^x, \quad x \in \mathbb{N}$$
where \( \alpha \in (0, 1) \) is the rate constant. When \( X \) has this geometric distribution, the partial sum process and the record process are the same, and the sequence \( \mathbf{X} = (X_1, X_2, \ldots) \) gives the total number of failures before the successes in the Bernoulli trials process with success parameter \( \alpha \). In particular for \( n \in \mathbb{N}_+ \), \( X_n \) has the negative binomial distribution with parameters \( \alpha \) and \( n \), with density function \( f_n \) defined by

\[
f_n(x) = \binom{x + n - 1}{n - 1} \alpha^n (1 - \alpha)^x, \quad x \in \mathbb{N}
\]

Moreover, the Bernoulli trials process is the “most random” way to put points in \( \mathbb{N} \) in the sense that given \( X_{n+1} = x \in \mathbb{N} \), the random vector \( (X_1, X_2, \ldots, X_n) \) is uniformly distributed on

\[
\{(x_1, x_2, \ldots, x_n) \in \mathbb{N}^n : x_1 \leq x_2 \leq \cdots \leq x_n \leq y\}
\]
a set that has measure (cardinality) \( \binom{x+n}{n} \). Note the repeated appearance of the expression \( \binom{x+k}{k} \) for various \( x \in \mathbb{N} \) and \( k \in \mathbb{N} \).

**The Abstract Setting**

Many of the results above generalize to a surprisingly abstract setting, and exploring this generalization is the main goals of this text. In this abstract setting,

- A measure space \((S, \mathcal{S}, \lambda)\) replaces \([0, \infty)\) with the Euclidean measure structure, or \(\mathbb{N}\) with the discrete measure structure.
- A (measurable) binary relation \(\rightarrow\) on \(S\) replaces the order relation \(\leq\).
- A (measurable) semigroup operator \(\cdot\) on \(S\) replaces the addition operator \(+\).

In terms of the measure structure, the only assumption that we need for the most part, is that the space has a measurable diagonal. Almost all of the basic theory goes through with just this assumption, and so we are able to stay out of the measure-theoretic and topological weeds.

The generalized “graph” \((S, \rightarrow)\) is the natural home for generalizations of concepts that rely only on the order relation \(\leq\). In particular, the reliability function is replaced by the right probability function \(F\) defined by \(F(x) = \mathbb{P}(x \rightarrow X)\) for \(x \in S\). Naturally, the right probability function does not in general uniquely determine the distribution, but we can give conditions under which it does. The rate function is given as before by \(f/F\), where \(f\) is the density function with respect to \(\lambda\), so in particular, constant rate distributions make sense. When the relation is a partial order, the concepts of increasing and decreasing rate make sense as well. Under mild conditions, generalizations of (1) and (3) hold with \(x \mapsto x^n/n!\) and \(x \mapsto \binom{n+x}{n}\) for \(n \in \mathbb{N}\) replaced by

\[
\gamma_n(x) = \lambda^n \{(x_1, x_2, \ldots, x_n) \in S^n : x_1 \rightarrow x_2 \rightarrow \cdots \rightarrow x_n \rightarrow x\}, \quad x \in S
\]

the measure of the set of “walks” in the graph of length \(n\) terminating in \(x\). The random walk associated with a distribution has transition density \(P\) given by \(P(x, y) = f(y)/F(x)\) when \(x \rightarrow y\), and can be constructed via record values. When a constant rate distribution exists, the corresponding random walk \((X_1, X_2, \ldots)\) is the “most random” way to put points in the graph, in the sense that given \(X_{n+1} = x \in S\), the random vector \((X_1, X_2, \ldots, X_n)\) is uniformly distributed on the set of walks of length \(n\) terminating in \(x\). Generalizations of (2) and (4) hold, so that if the rate constant is \(\alpha \in (0, \infty)\) then \(X_n\) has density function \(f_n\) given by

\[
f_n(x) = \alpha^n \gamma_{n-1}(x)F(x), \quad x \in S
\]

A finite, strongly connected graph has a unique constant rate distribution, a result that turns out to be the Perron-Frobenius theorem in disguise.

A semigroup \((S, \cdot)\) satisfying the left cancellation property is the natural home for (left) invariance of the reference measure \(\lambda\) so that \(\lambda(xA) = \lambda(A)\) for \(x \in S\) and \(A \in \mathcal{S}\). A semigroup has a natural relation \(\rightarrow\) associated with it, defined by \(x \rightarrow y\) if and only if there exists (a unique) \(t \in S\) with \(xt = y\), and then \(t\) is denoted \(x^{-1}y\). The brief sketch in the last paragraph applies to the graph \((S, \rightarrow)\), of course, but in addition, we can study concepts that involve the comparison of a shifted probability distribution to the original distribution. Random variable \(X\) has an exponential distribution if the conditional distribution of \(x^{-1}X\) given \(x \rightarrow X\) is the same as the distribution of \(X\). The generally weaker memoryless property just
requires the right probability function of \( x^{-1}X \) given \( x \to X \) be the same as the right probability function of \( X \). Under mild conditions, a distribution is exponential if and only if it is memoryless and has constant rate. With slightly more restrictions, the memoryless property implies the full exponential property. Given a probability distribution on \( S \) with density \( f \), the random walk on \((S, \to)\) in the last paragraph makes sense, but there is another random walk on \((S, \cdot)\) with transition density \( Q \) given by \( Q(x, y) = f(x^{-1}y) \) when \( x \to y \).

This random walk corresponds to the partial product sequence associated with a sequence of independent variables, each with density \( f \). The two random walks are the same if and only if the underlying distribution is exponential.

Much of the underlying mathematics in the general theory is simple and well known, corresponding to basic results in measure theory, linear algebra, functional analysis, and graph theory. But the application of the mathematics to the particular topics in probability theory presented here is not well known, to the best of my knowledge. But it should be. The theory is elegant and the applications interesting and diverse, even if they sometimes have little relation to the classical reliability theory that served as motivation.

Here are a few random examples:

1. As noted above, the geometric distribution is the exponential distribution for the standard discrete semigroup \((\mathbb{N}, +)\) and hence has constant rate for the graph \((\mathbb{N}, \leq)\). But it also has constant rate (with different constants) for the graphs \((\mathbb{N}, <)\) and \((\mathbb{N}, \rightarrow)\) where \( \rightarrow \) is the covering relation of \((\mathbb{N}, \leq)\), and for the graph \((\mathbb{N}, \Rightarrow)\) obtained by adding a loop to \((\mathbb{N}, \rightarrow)\) at each vertex. This is a pattern that repeats for other partial order graphs.

2. The unique constant rate distribution for a simple, undirected, finite path is a discrete version of Gilbert’s sine distribution. It seems surprising that a sine distribution would govern the most random way to put points in a path, but seen through the lens of spectral graph theory, the result is much clearer. So often, other areas of mathematics inform the theory here.

3. The exponential distributions for the standard arithmetic semigroup \((\mathbb{N}_+, \cdot)\) are the Dirichlet distributions corresponding to completely multiplicative functions. Some known properties of the Dirichlet distributions that seem rather mysterious in the context of analytic number theory are much clearer in the context of exponential distributions. So, conversely to item 2, sometimes the theory presented here informs other areas of mathematics.

4. For \( n, k \in \mathbb{N}_+ \), the relation \( \rightarrow_k \) associated with the \( k \) norm on \( \mathbb{R}^n \) is given by \( x \rightarrow_k y \) if and only if \( \|x\|_k \leq \|y\|_k \) for \( x, y \in \mathbb{R}^n \). The constant rate distributions for the graph \((\mathbb{R}^n, \rightarrow_k)\) form a class of generalized normal distributions. A random vector in \( \mathbb{R}^n \) with a constant rate distribution has independent components only in the case that \( k = n \) and in this case, the distributions reduce to a known class of generalized normal distribution. In particular, the Laplace distribution has constant rate when \( n = k = 1 \) and a class of ordinary normal distribution have constant rate when \( n = k = 2 \).

### Structure of the Text

The text is divided into three basic parts and a couple of appendices. Part I gives the general theory: basic definitions and properties of graphs and semigroups, probability distributions and random walks on graphs and semigroups, and finally exponential, memoryless, and constant rate distributions. Part II concerns basic constructions for creating new graphs and semigroups from existing ones. These include various product structures, a quotient structure, and graphs induced by an underlying graph and a measurable function. Part III studies a number of applications and examples in detail and is perhaps the most interesting part. The applications include the standard discrete and continuous spaces (and derivatives), rooted trees, the free semigroup on a countable alphabet, arithmetic semigroups including the standard \((\mathbb{N}_+, \cdot)\), and the subset partial order.

My hope is that this text will be interesting and useful to students and researchers who study the interplay between probability and algebraic structures. The prerequisites are measure theory, probability and stochastic processes, linear algebra, and graph theory at the advanced undergraduate or beginning graduate levels. With students in mind, I have sprinkled the text with a number of simple exercises. Solutions to the exercises are given in Appendix B. The \( \checkmark \) symbol at the end of an exercise is a hyperlink to the answer to that exercise. With researchers in mind, I have also included a number of problems that are interesting to me, but whose
solutions I do not know. Hyperlinks are given to basic topics in probability, measure theory, and stochastic processes in the interactive web text Random. As illustrated, these hyperlinks are in red.

This book is currently a work in progress, and may well contain mistakes, both minor and serious. It will be updated regularly at least until it becomes reasonably correct and stable. I am grateful for comments and corrections.
Contents

I General Theory

1 Measure
   1.1 Measure Spaces ................................................................. 2
   1.2 Kernels ........................................................................ 3

2 Graphs
   2.1 Basics ................................................................. 5
   2.2 Partial Orders ........................................................... 11
   2.3 Möbius Inversion ............................................................ 13

3 Semigroups
   3.1 Basics ................................................................. 15
   3.2 Positive Semigroups ....................................................... 19
   3.3 Left Invariance ............................................................. 21

4 Probability on Graphs
   4.1 Basics ................................................................. 24
   4.2 Moments ................................................................. 27
   4.3 Random Walks ............................................................ 30
   4.4 Constant Rate Distributions ............................................. 33

5 Probability on Semigroups
   5.1 Basics ................................................................. 41
   5.2 Random Walks ............................................................ 42
   5.3 Exponential and Memoryless Distributions ............................. 44
   5.4 Conditional Distributions ............................................... 50

II Special Constructions

6 Reflexive Completion ............................................................. 55

7 Induced Graphs
   7.1 Introduction .............................................................. 59
   7.2 Distributions ............................................................. 61
   7.3 Random Walks ............................................................. 63
   7.4 The Discrete Case .......................................................... 65
      7.4.1 Equivalence Relations .............................................. 66
      7.4.2 Complete Multipartite Graphs .................................... 68

8 Product Spaces
   8.1 Direct Product of Graphs ............................................... 71
   8.2 Cartesian Products of Graphs .......................................... 73
   8.3 Direct Products of Semigroups ......................................... 74
   8.4 Lexicographic Sums of Graphs ......................................... 78
## Contents

### 9 Quotient spaces
- 9.1 Preliminaries .............................................. 83
- 9.2 Distributions ............................................... 85

### III Examples and Applications

#### 10 Standard Continuous Spaces
- 10.1 The standard space ....................................... 89
- 10.2 Spaces isomorphic to the standard space
  - 10.2.1 The shifted space .................................. 91
  - 10.2.2 The Beta Distribution .............................. 94
  - 10.2.3 The Pareto Distribution ............................ 95
  - 10.2.4 An application to Brownian functionals .......... 95
- 10.3 Applications to reliability
  - 10.3.1 Examples ........................................... 98
  - 10.3.2 Strong aging properties ......................... 99
- 10.4 Norm Graphs .............................................. 100

#### 11 Standard Discrete Spaces
- 11.1 Basics .................................................... 106
- 11.2 Distributions ............................................. 107
- 11.3 Random Walks .......................................... 110
- 11.4 Quotient spaces ........................................ 112

#### 12 Rooted Trees
- 12.1 Basics .................................................... 114
- 12.2 Probability .............................................. 115

#### 13 The Free Semigroup
- 13.1 Basics .................................................... 119
- 13.2 Probability .............................................. 121

#### 14 Arithmetic Semigroups
- 14.1 Basics .................................................... 126
- 14.2 Probability .............................................. 131
  - 14.2.1 Distributions in General ......................... 131
  - 14.2.2 Exponential Distributions ....................... 132
  - 14.2.3 Random Walks ..................................... 137
  - 14.2.4 Related Graphs ................................... 138

#### 15 Paths and Grids
- 15.1 Preliminaries ............................................ 140
- 15.2 The Infinite Path ....................................... 141
- 15.3 Finite Paths ............................................. 144
- 15.4 Grid Graphs ............................................. 148

#### 16 The Subset Semigroup
- 16.1 Preliminaries ............................................ 149
- 16.2 The Positive Semigroup .............................. 151
- 16.3 Sub-semigroups .......................................... 155
- 16.4 Exponential distributions ............................ 159
- 16.5 Almost exponential distributions ................... 162
- 16.6 Constant rate distributions ......................... 169
CONTENTS

17 Induced Graphs 175
  17.1 The Cover Graph .................................................. 175
  17.2 The Completed Cover Graph ..................................... 178
  17.3 The Strict Partial Order Graph ................................. 179
  17.4 The Partial Order Graph ......................................... 180
  17.5 The Lexicographic Sum ........................................... 180
  17.6 Distributions ...................................................... 182
  17.7 Positive semigroups ............................................... 184
  17.8 Quotient Spaces .................................................. 186

18 Miscellaneous Examples 188
  18.1 Remaining Life Chain ............................................. 188
  18.2 The Success Runs Chain ......................................... 189
  18.3 The Wheel .......................................................... 191

IV Appendices 194
  A Characterizations of the Poisson Distribution 195
  B Solutions to Exercises 199
Part I

General Theory
Chapter 1
Measure

1.1 Measure Spaces

This chapter reviews the basics of measure theory that we will need. Recall that a measurable space \((S, \mathcal{F})\) consists of a set \(S\) and a \(\sigma\)-algebra \(\mathcal{F}\) of subsets of \(S\). If \(\lambda\) is a positive measure on \((S, \mathcal{F})\) then \((S, \mathcal{F}, \lambda)\) is a measure space. The measure space is \(\sigma\)-finite if there exists a countable collection \(\{A_i : i \in I\}\) of sets in \(\mathcal{F}\) that partition \(S\) with \(\lambda(A_i) < \infty\) for \(i \in I\). If \(A \in \mathcal{F}\) then \(1_A\) denotes the indicator function of \(A\) on \(S\). That is, \(1_A(x) = 1\) if \(x \in A\) and \(1_A(x) = 0\) if \(x \in S - A\). In particular, \(1_S\) is the constant function 1 on \(S\), denoted simply by 1 when \(S\) is understood.

If \((S, \mathcal{F})\) and \((T, \mathcal{G})\) are measurable spaces, then \(\mathcal{F} \times \mathcal{G}\) denotes the product \(\sigma\)-algebra on the Cartesian product space \(S \times T\) generated by products sets of the form \(A \times B\) where \(A \in \mathcal{F}\) and \(B \in \mathcal{G}\). If \(\lambda\) and \(\mu\) are \(\sigma\)-finite measures for \((S, \mathcal{F})\) and \((T, \mathcal{G})\) respectively, then the corresponding product measure \(\lambda \times \mu\) for \((S \times T, \mathcal{F} \times \mathcal{G})\) is the unique positive measure (also \(\sigma\)-finite) that satisfies

\[(\lambda \times \mu)(A \times B) = \lambda(A)\mu(B), \quad A \in \mathcal{F}, B \in \mathcal{G}\]

The product construction extends in the obvious way to a product of \(n\) measurable spaces \((S_1, \mathcal{F}_i, \lambda_i)\) for \(i \in \{1, 2, \ldots, n\}\) and \(n \in \mathbb{N}_+\). In particular, if \((S, \mathcal{F}, \lambda)\) is a basic measure space then \((S^n, \mathcal{F}^n, \lambda^n)\) is the corresponding \(n\)-fold power space for \(n \in \mathbb{N}_+\).

**Definition 1.1.** A measurable space \((S, \mathcal{F})\) has a measurable diagonal if

\[D = \{(x, x) : x \in S\} \in \mathcal{F}^2\]

We will assume that the measurable spaces considered in this text for graphs and semigroups have measurable diagonals. In this case, it follows that \(\{x\} \in \mathcal{F}\) for \(x \in S\), and hence \(\mathcal{F}\) contains all countable subsets of \(S\). The assumption gives other benefits as well, as we will see later.

**Proposition 1.1.** If \((S, \mathcal{F})\) and \((T, \mathcal{G})\) are measurable spaces with measurable diagonals then \((S \times T, \mathcal{F} \times \mathcal{G})\) also has a measurable diagonal.

**Proof.** The mapping \(\varphi : (S \times T)^2 \to S^2 \times T^2\) given by

\[\varphi([x_1, y_1], [x_2, y_2]) = [(x_1, x_2), (y_1, y_2)], \quad x_1, x_2 \in S, y_1, y_2 \in T\]

is one-to-one, onto, and measurable with respect to the appropriate product \(\sigma\)-algebras \((\mathcal{F} \times \mathcal{G})^2\) and \(\mathcal{F}^2 \times \mathcal{G}^2\). Let \(D_S\), \(D_T\), and \(D_{S \times T}\) denote the diagonals for the sets \(S\), \(T\), and \(S \times T\), respectively. By assumption, \(D_S \in \mathcal{F}^2\), \(D_T \in \mathcal{G}^2\) and hence \(S_S \times D_T \in \mathcal{F}^2 \times \mathcal{G}^2\). Therefore

\[D_{S \times T} = \varphi^{-1}(D_S \times D_T) \in (\mathcal{F} \times \mathcal{G})^2\]

\[\square\]

If \((S, \mathcal{F})\) is a measurable space, then we usually have a fixed \(\sigma\)-finite measure \(\lambda\) for \((S, \mathcal{F})\), referred to as a reference measure. Discrete spaces will be particularly important in this study.
Definition 1.2. The measurable space \((S, \mathcal{S})\) is **discrete** if \(S\) is countable and \(\mathcal{S}\) is the \(\sigma\)-algebra of all subsets of \(S\). Counting measure \(#\) is the reference measure on \((S, \mathcal{S})\).

Note that a discrete space has a measurable diagonal and is \(\sigma\)-finite. A product of a finite number of discrete spaces is also discrete.

The product construction above also extends to an infinite collection of measure spaces. If \((S_i, \mathcal{S}_i, \lambda_i)\) is a measure space for \(i\) in a nonempty index set \(I\) then \(\prod_{i \in I} \mathcal{S}_i\) is the \(\sigma\)-algebra of subsets of \(\prod_{i \in I} S_i\) generated by **cylinder sets** of the form \(\prod_{i \in I} A_i\), where \(A_i \in \mathcal{S}_i\) for \(i \in I\) and \(A_i = S_i\) for all but finitely many \(i \in I\). The corresponding product measure \(\lambda := \prod_{i \in I} \lambda_i\) is the unique positive measure on \((\prod_{i \in I} S_i, \prod_{i \in I} \mathcal{S}_i)\) that satisfies

\[
\lambda \left( \prod_{i \in I} A_i \right) = \prod_{i \in I} \lambda_i(A_i)
\]

where \(\prod_{i \in I} A_i\) is a cylinder set.

If \((S, \mathcal{S}, \lambda)\) is a measure space, then usually the set \(S\) will have a nice topology, for example, a locally compact topology with a countable base (LCCB). Then \(\mathcal{S}\) will be the Borel \(\sigma\)-algebra (the \(\sigma\)-algebra generated by the topology), and \(\lambda\) will be a Borel measure (a positive measure satisfying \(\lambda(C) < \infty\) if \(C \subseteq S\) is compact). In particular, the Borel measure space of an LCCB has a measurable diagonal. However, our goal is to develop the basic theory with minimal assumptions, so we will not impose a particular topological structure unless necessary.

If \((S, \mathcal{S})\) is a measurable space then \(\mathcal{M}\) will denote the collection of measurable functions from \(S\) to \(\mathbb{R}\) and \(\mathcal{M}_+ = \{f \in \mathcal{M} : f \geq 0\}\), the set of nonnegative measurable functions. If \((S, \mathcal{S}, \lambda)\) is a measure space and \(k \in \mathbb{N}_+\), then the usual Lebesgue function space is denoted \(L_k\). This is the space functions \(f \in \mathcal{M}\) with \(\|f\|_k < \infty\) where

\[
\|f\|_k^2 = \int |f(x)|^k \, d\lambda(x) < \infty
\]

1.2 **Kernels**

For the following definitions, we assume that we have a fixed \(\sigma\)-finite measure space \((S, \mathcal{S}, \lambda)\).

**Definition 1.3.** A kernel on \(S\) is a measurable function \(K : S^2 \to \mathbb{R}\). A kernel \(K\) defines two operators on measurable functions \(f : S \to \mathbb{R}\), one on the left and one on the right, assuming that the integrals exist:

\[
(Kf)(x) = \int_S K(x, y) f(y) \, d\lambda(y), \quad x \in S
\]

\[
(fK)(y) = \int_S f(x) K(x, y) \, d\lambda(x), \quad y \in S
\]

If \(S\) is finite, a kernel function is simply a square matrix, with rows and columns indexed by the elements of \(S\). The operations on a function are ordinary multiplication of the matrix by a column vector on the right, or a row vector on the left, again with the entries of the vector indexed by the elements of \(S\). So in general, one can think of kernel as the natural abstraction of a matrix.

Suppose that \(K\) and \(L\) are kernels on \(S\) and that \(a \in \mathbb{R}\). Then \(K + L\) and \(aK\) are also kernels with the usual pointwise definitions:

\[
(K + L)(x, y) = K(x, y) + L(x, y), \quad (x, y) \in S^2
\]

\[
(aK)(x, y) = aK(x, y), \quad (x, y) \in S^2
\]

If \(S\) is finite these operations are the ordinary addition and scalar multiplication for matrices. The kernel \(0\) defined by \(0(x, y) = 0\) for \((x, y) \in S^2\) is clearly the additive identity, so the collection \(\mathcal{K}\) of kernels on \(S\) is a vector space. Multiplication is not quite as simple, but is also very natural.

**Definition 1.4.** If \(K\) and \(L\) are kernels on \(S\) then \(KL\) is the kernel defined as follows, assuming that the integrals exist in \(\mathbb{R}\).

\[
(KL)(x, y) = \int_S K(x, t) L(t, y) \, d\lambda(t), \quad (x, y) \in S^2
\]
Again, if $S$ is finite, then $KL$ is just ordinary matrix multiplication. The usual associative and distributive properties hold. So if $K, L, M$ are kernels on $S$, then assuming that the appropriate integrals exist,

$$(KL)M = K(LM), \quad K(L + M) = KL + KM, \quad (K + L)M = KM + LM$$

Define the function $I$ by $I(x, y) = 1(x = y)$ for $(x, y) \in S^2$. Note that since $(S, \mathcal{F})$ has a measurable diagonal, $I$ is measurable and hence is a kernel on $S$. In the discrete case, $I$ is the identity with respect to the multiplication operation, so that $IK = KI = K$ for every kernel $K$ on $S$. Finally, if $K$ is a kernel then $K^n$ denotes the $n$-fold power for $n \in \mathbb{N}$, with the convention that $K^0 = I$. 
Chapter 2

Graphs

This chapter is devoted to the algebraic and measure-theoretic structure of graphs, one of the two basic spaces that we study in this text.

2.1 Basics

Our underlying space is a \( \sigma \)-finite measure space \((S, \mathcal{S}, \lambda)\) with a measurable diagonal. In the discrete space, as usual, the reference measure is counting measure \(#\).

Definition 2.1. A graph \((S, \rightarrow)\) consists of the set \(S\) and a measurable binary relation \(\rightarrow\) on \(S\). That is,

\[ \{(x, y) \in S^2 : x \rightarrow y\} \in \mathcal{S}^2 \]

Technically, the relation \(\rightarrow\) is the set of ordered pairs above, but we will use notation more appropriate for relations rather than sets. In the discrete case, \((S, \rightarrow)\) is a graph in the usual combinatorial sense, with \(S\) as the set of vertices and \(\rightarrow\) as the set of edges (so in general, such a graph may be directed and may have loops). Keeping with graph terminology, when \(x \rightarrow y\) we say that \((x, y)\) is an edge in the graph. Note that the assumption of a measurable diagonal simply means that \((S, \rightarrow)\) is a graph, clearly an essential requirement. The following definition reviews some standard properties of relations.

Definition 2.2. A graph \((S, \rightarrow)\) is

(a) Reflexive if \(x \rightarrow x\) for all \(x \in S\).

(b) Irreflexive if no \(x \in S\) satisfies \(x \rightarrow x\).

(c) Symmetric if \(x \rightarrow y\) implies \(y \rightarrow x\) for all \(x, y \in S\).

(d) Asymmetric if no \(x, y \in S\) satisfies \(x \rightarrow y\) and \(y \rightarrow x\).

(e) Anti-symmetric if \(x \rightarrow y\) and \(y \rightarrow x\) imply \(x = y\) for all \(x, y \in S\).

(f) Transitive if \(x \rightarrow y\) and \(y \rightarrow z\) imply \(x \rightarrow z\) for all \(x, y, z \in S\).

We use the same terminology for the relation \(\rightarrow\). So in the discrete case, a symmetric, irreflexive graph is an unordered graph in the usual combinatorial sense. A reflexive, antisymmetric, transitive relation is a partial order. A reflexive, symmetric, transitive relation is an equivalence relation. These special relations will be studied in more detail in later sections. We can form new graphs from given ones using the standard set operations. The following definitions give the most important cases of this.

Definition 2.3. Suppose that \((S, \rightarrow)\) is a graph. If \(R\) is a measurable subset of \(S\), and \(\uparrow\) is a measurable relation on \(R\) with the property that \(x \uparrow y\) implies \(x \rightarrow y\) for \(x, y \in R\) then \((R, \uparrow)\) is a subgraph of \((S, \rightarrow)\).

So to rephrase, \(R\) is a measurable subset of \(S\) and as sets of ordered pairs, \(\uparrow\) is a measurable subset of \(\rightarrow\). In the discrete case, subgraph has its usual meaning: the vertices and edges in the subgraph are also vertices and edges in the parent graph.
Proposition 2.1. Suppose that $(S, \rightarrow)$ is a graph. Then the following are also graphs:
(a) $(S, \leftarrow)$ where $\leftarrow$ is the reverse of $\rightarrow$, so that $x \leftarrow y$ if and only if $y \rightarrow x$.
(b) $(S, \not\rightarrow)$ where $\not\rightarrow$ is the complement of $\rightarrow$, so that $x \not\rightarrow y$ if and only if it is not true that $x \rightarrow y$.

Exercise 2.1. Prove Proposition 2.1.

Proposition 2.2. Suppose that $(S, \rightarrow)$ and $(S, \uparrow)$ are graphs. Then the following are also graphs:
(a) $(S, \cup)$ where $\cup$ is the union of $\rightarrow$ and $\uparrow$, so that $x \cup y$ if and only if $x \rightarrow y$ or $x \uparrow y$.
(b) $(S, \cap)$ where $\cap$ is the intersection of $\rightarrow$ and $\uparrow$, so that $x \cap y$ if and only if $x \rightarrow y$ and $x \uparrow y$.
(c) $(S, \setminus)$ where $\setminus$ is the set difference of $\rightarrow$ and $\uparrow$, so that $x \setminus y$ if and only if $x \rightarrow y$ but not $x \uparrow y$.

Exercise 2.2. Prove Proposition 2.2.

Proposition 2.3. Suppose again that $(S, \rightarrow)$ and $(S, \uparrow)$ are graphs. Then $(S, \Rightarrow)$ is also a graph, where $\Rightarrow$ is the composition of $\rightarrow$ with $\uparrow$, so that $x \Rightarrow y$ if and only if there exists $z \in S$ with $x \rightarrow z$ and $z \uparrow y$.

Proof. By assumption $\{(x, u) \in S^2 : x \rightarrow u\} \in \mathcal{P}^2$ and $\{(v, y) \in S^2 : v \uparrow y\} \in \mathcal{P}^2$. Hence
\[
\{(x, u, v, y) \in S^4 : x \rightarrow u, v \uparrow y\} \in \mathcal{P}^4
\]
By the assumption of a measurable diagonal, $\{(x, z, z, y) : x, y \in S\} \in \mathcal{P}^4$. Intersecting the two sets we have
\[
\{(x, z, z, y) \in S^4 : x \rightarrow z, z \uparrow y\} \in \mathcal{P}^4
\]
The projection of this set onto the first and last coordinates is measurable. That is,
\[
\{(x, y) \in S^2 : x \Rightarrow y\} = \{(x, y) \in S^2 : \text{there exists } z \in S \text{ with } x \rightarrow z, z \uparrow y\} \in \mathcal{P}^2
\]

Proposition 2.4. Properties of composition.

(a) Composition is associative.
(b) The equality relation $=$ is the identity with respect to composition.

Proof. For the proofs, we use product notation for composition.

(a) Suppose that $\rightarrow_1, \rightarrow_2, \rightarrow_3$ are relations on $S$. From the definition, it’s straightforward to see that $x \rightarrow_1 (\rightarrow_2 \rightarrow_3)[y]$ if and only if $x \rightarrow_1 (\rightarrow_2 \rightarrow_3)[y]$ if and only if there exist $u, v \in S$ with $x \rightarrow_1 u, u \rightarrow_2 v,$ and $v \rightarrow_3 y$.

(b) Suppose that $\rightarrow$ is a relation on $S$. Then $x = \rightarrow y$ if and only if there exists $u \in S$ with $x = u$ and $u \rightarrow y$ if and only if $x \rightarrow y$. Similarly, $x = \rightarrow y$ if and only if $x \rightarrow y$.

Definition 2.4. Suppose that $(S, \rightarrow)$ is a graph and that $n \in \mathbb{N}_+$ and $(x_1, x_2, \ldots , x_{n+1}) \in S^{n+1}$. The notation
\[
x_1 \rightarrow x_2 \rightarrow \cdots \rightarrow x_{n+1}
\]
means that $x_i \rightarrow x_{i+1}$ for each $i \in \{1, 2, \ldots , n\}$. The sequence is a walk of length $n$ in the graph from $x_1$ to $x_{n+1}$. If the vertices are distinct, the sequence is a path of length $n$ from $x_1$ to $x_{n+1}$.

In the notation above, it’s important to keep in mind that the relation $\rightarrow$ is not necessarily transitive. So, for example, $x \rightarrow y \rightarrow z$ does not imply $x \rightarrow z$. In the discrete setting, the terms walk and path have their usual meanings: a walk is sequence of adjacent vertices in the graph while a path is a sequence of distinct adjacent vertices in the graph. For $n \in \mathbb{N}_+$, let $\rightarrow^n$ denote the $n$-fold composition power of $\rightarrow$. Note that $x \rightarrow^n y$ if and only if there exists a walk for length $n$ from $x$ to $y$. 
Going forward, we will define a number of algebraic and probabilistic objects for a graph \((S, \rightarrow)\). In general, these objects come in pairs—a left object defined in terms of the mapping that takes \(x\) to \(\{u \in S : u \rightarrow x\}\) (the set of points that are related to \(x\)) and a right object defined in terms of the mapping that takes \(x\) to \(\{y \in S : x \rightarrow y\}\) (the set of points that \(x\) is related to). All right objects of \((S, \rightarrow)\) are left objects of the reverse graph \((S, \leftarrow)\). Since our relation \(\rightarrow\) is arbitrary, there would be no loss in generality in defining just one type of object. However, it is often the relationship between the left and right objects that is the most interesting. So our primary definitions, which we will give explicitly, will sometimes be left objects and sometimes right object, depending on the context. Of course, if \(\rightarrow\) is symmetric, the right and left objects are the same and we can drop the adjective.

**Definition 2.5.** A kernel of the graph \((S, \rightarrow)\) is a measurable function \(K : S^2 \rightarrow \mathbb{R}\) with the property that \(K(x, y) = 0\) if \(x \not\rightarrow y\). The adjacency kernel \(L\) of \((S, \rightarrow)\) is the indicator function of \(\rightarrow\) (as a set of ordered pairs), so that \(L(x, y) = 1(x \rightarrow y)\) for \((x, y) \in S^2\).

So a kernel on the graph \((S, \rightarrow)\) is a kernel on \(S\) as defined earlier, but with the additional property that the kernel respects the relation \(\rightarrow\). Since the relation \(\rightarrow\) is measurable, so is the adjacency kernel \(L\) and hence the definition makes sense. In the discrete case, \(L\) is the adjacency matrix of the graph. Various collections of walks are measurable:

**Theorem 2.1.** Suppose that \((S, \rightarrow)\) is a graph, and let \(n \in \mathbb{N}_+\) and \(x, y \in S\). Then

(a) \(\{(x_1, x_2, \ldots, x_{n+1}) \in S^{n+1} : x_1 \rightarrow x_2 \rightarrow \cdots \rightarrow x_{n+1}\} \in \mathcal{J}^{n+1}\).

(b) \(\{(x_1, x_2, \ldots, x_n) \in S^n : x \rightarrow x_1 \rightarrow \cdots \rightarrow x_n\} \in \mathcal{J}^n\).

(c) \(\{(x_1, x_2, \ldots, x_n) \in S^n : x_1 \rightarrow \cdots \rightarrow x_n \rightarrow y\} \in \mathcal{J}^n\).

(d) \(\{(x_1, x_2, \ldots, x_n) \in S^n : x \rightarrow x_1 \rightarrow \cdots \rightarrow x_n \rightarrow y\} \in \mathcal{J}^n\).

**Proof.** Let \(L\) denote the adjacency kernel of \((S, \rightarrow)\) and let \(n \in \mathbb{N}_+\). Recall that cross sections of measurable sets are measurable by standard results in measure theory.

(a) This is the set of all walks of length \(n\) in \(S\). The indicator function of this set is the function on \(S^{n+1}\) given by

\[
(x_1, x_2, \ldots, x_{n+1}) \mapsto L(x_1, x_2)L(x_2, x_3) \cdots L(x_n, x_{n+1})
\]

which is clearly measurable.

(b) For \(x \in S\), this is the cross section on the left at \(x\) of the set of walks of length \(n\).

(c) For \(y \in S\), this is the cross section on the right at \(y\) of the set of walks of length \(n\).

(d) For \(x, y \in S\), \((x, y) \in S^2\), this the cross section at \(x\) on the left and \(y\) on the right of the set of walks of length \(n + 1\).

If \(K\) is a kernel for the graph \((S, \rightarrow)\) then the domains of integration for the operators associated with \(K\) are reduced. Specifically, if \(f : S \rightarrow \mathbb{R}\) is measurable then

\[
(Kf)(x) = \int_{x\rightarrow y} K(x, y)f(y)\,d\lambda(y), \quad x \in S
\]

\[
(fK)(y) = \int_{x\rightarrow y} f(x)K(x, y)\,d\lambda(x), \quad y \in S
\]

assuming as usual that the integrals exist. In particular, if \(L\) is the adjacency kernel then for \(f \in \mathcal{M}_+\) or if \(f \in \mathcal{A}_1\),

\[
(Lf)(x) = \int_{x\rightarrow y} f(y)\,d\lambda(y), \quad x \in S
\]

\[
(fL)(y) = \int_{x\rightarrow y} f(x)\,d\lambda(x), \quad y \in S
\]
For \( f, g \in M_+ \) or for \( f, g \in \mathcal{L}_2 \), the self-adjoint property of the adjacency kernel \( L \) is

\[
(fL, g) = (f, Lg) = \int_{x\to y} f(x)g(y) \, d\lambda^2(x, y)
\]

If \( J, K \) are kernels of \( (S, \to) \) then

\[
(JK)(x, y) = \int_{x\to t\to y} J(x, t)K(t, y) \, d\lambda(t), \quad (x, y) \in S^2
\]

Clearly if \( J, K \) are kernels of \( (S, \to) \) then so are \( J + K \) and \( aK \) for \( a \in \mathbb{R} \). On the other hand, \( JK \) is a kernel of \( (S, \to^2) \), where \( \to^2 \) is the composition power of \( \to \) of order 2. Powers of the adjacency kernel give information about the measure of walks in the graph.

**Theorem 2.2.** Suppose that \( (S, \to) \) is a graph with adjacency kernel \( L \).

(a) For \( n \in \{2, 3, \ldots\} \),

\[
L^n(x, y) = \lambda^{n-1}\{(x_1, x_2, \ldots, x_{n-1}) \in S^{n-1} : x \to x_1 \to \cdots \to x_{n-1} \to y\}, \quad (x, y) \in S^2
\]

(b) For \( n \in \mathbb{N}_+ \),

\[
\begin{align*}
(1L^n)(x) &= \lambda^n\{(x_1, \ldots, x_n) \in S^n : x_1 \to \cdots \to x_n \to x\}, \quad x \in S \\
(L^n1)(x) &= \lambda^n\{(x_1, \ldots, x_n) \in S^n : x \to x_1 \to \cdots \to x_n\}, \quad x \in S
\end{align*}
\]

So if \( n \in \{2, 3, \ldots\} \) then \( L^n(x, y) \) is the measure of the set of the middle parts of walks of length \( n \) from \( x \) to \( y \). If \( n \in \mathbb{N}_+ \) then \( (1L^n)(x) \) is the measure of the set of the initial parts of paths of length \( n + 1 \) that terminate in \( x \), while \( (L^n1)(x) \) is the measure of the terminal parts of paths of length \( n + 1 \) that start in \( x \). The first of these will play an important role.

**Definition 2.6.** Suppose again that \( (S, \to) \) is a graph with adjacency kernel \( L \). For \( n \in \mathbb{N} \), the left walk function of order \( n \) for \( (S, \to) \) is the function \( w_n = 1L^n : S \to [0, \infty] \), so that \( w_0(x) = 1 \) for \( x \in S \), and for \( n \in \mathbb{N}_+ \),

\[
w_n(x) = \lambda^n\{(x_1, \ldots, x_n) \in S^n : x_1 \to \cdots \to x_n \to x\}
\]

We often abbreviate \( w_1 \) by \( w \) so that \( w_1(x) = \lambda\{u \in S : u \to x\} \) for \( x \in S \), the measure of the set of elements related to \( x \). To rephrase the definition, the left walk functions are defined recursively by \( w_0(x) = 1 \) for \( x \in S \) and for \( n \in \mathbb{N} \),

\[
w_{n+1}(x) = \int_{y\to x} w_n(y), \quad x \in S
\]

**Definition 2.7.** A graph \( (S, \to) \) with adjacency kernel \( L \) is strongly connected if for every \( (x, y) \in S^2 \), there exists \( n \in \mathbb{N}_+ \) with \( L^n(x, y) > 0 \).

In the discrete case, this has its usual meaning: \( (S, \to) \) is strongly connected if there exists a (directed) walk between any two vertices.

**Definition 2.8.** A graph \( (S, \to) \) is left finite if \( w(x) = \lambda\{u \in S : u \to x\} < \infty \) for almost every \( x \in S \).

**Proposition 2.5.** Suppose that \( (S, \to) \) is a transitive graph. Then for \( n \in \mathbb{N} \), \( w_n(x) \leq w^n(x) \) for almost all \( x \in S \).

**Proof.** The statement is true for \( n = 0 \) by definition. Suppose the statement is true for a given \( n \in \mathbb{N} \). Since \( \to \) is transitive, \( y \to x \) implies \( \{u \in S : u \to y\} \subseteq \{u \in S : u \to x\} \) and so \( w(y) \leq w(x) \) almost always when \( y \to x \). Hence by the induction hypothesis, for almost all \( x \in S \),

\[
w_{n+1}(x) = \int_{y\to x} w_n(y) \, d\lambda(y) \leq \int_{y\to x} w^n(y) \, d\lambda(y) \leq \int_{y\to x} w^n(x) \, d\lambda(y) = w^{n+1}(x)
\]

\qed
CHAPTER 2. GRAPHS

So for a transitive, left-finite graph, \( w_n(x) < \infty \) for almost all \( x \in S \) and all \( n \in \mathbb{N} \).

**Definition 2.9.** For \( k \in (0, \infty) \), the graph \((S, \rightarrow)\) is right \( k \)-regular if \( \lambda\{y \in S : x \rightarrow y\} = k \) for all \( x \in S \).

This has its usual meaning for discrete, symmetric graphs: for \( k \in \mathbb{N}_+ \), the graph is \( k \)-regular if every vertex has \( k \) neighbors. Our last definition in this section is the left generating function associated with the sequence of left walk functions.

**Definition 2.10.** The left generating function \( W \) of the graph \((S, \rightarrow)\) is defined by

\[
W(x, t) = \sum_{n=0}^{\infty} w_n(x) t^n; \quad x \in S, \ |t| < \rho(x)
\]

where \( \rho(x) \) is the radius of convergence of the power series at \( x \).

For fixed \( x \in S \), \( t \mapsto W(x, t) \) is the generating function (in the combinatorial sense) of the sequence \((w_n(x) : n \in \mathbb{N})\). So, assuming that \( \rho(x) > 0 \), \( t \mapsto W(x, t) \) determines \((w_n(x) : n \in \mathbb{N})\). The generating function satisfies a simple integral equation.

**Proposition 2.6.** Suppose that \((S, \rightarrow)\) is a graph with left generating function \( W \). Then

\[
W(x, t) = 1 + t \int_{y \rightarrow x} W(y, t) \, d\lambda(y), \quad x \in S, \ |t| \leq \rho(x)
\]

**Proof.** The result follows easily from the recursive definition of the walk functions. For \( x \in S \) and \( |t| < \rho(x) \),

\[
1 + t \int_{y \rightarrow x} W(y, t) \, d\lambda(y) = 1 + t \int_{y \rightarrow x} \sum_{n=0}^{\infty} w_n(y) t^n \, d\lambda(y)
\]

\[
= 1 + t \sum_{n=0}^{\infty} t^n \int_{y \rightarrow x} w_n(y) \, d\lambda(y) = 1 + \sum_{n=0}^{\infty} w_{n+1}(x) t^{n+1} = W(x, t)
\]

\( \square \)

**Example 2.1.** Consider the standard continuous graph \(((0, \infty), \leq)\).

(a) For \( n \in \mathbb{N} \),

\[
w_n(x) = \frac{x^n}{n!}, \quad x \in [0, \infty)
\]

(b) For \( n \in \mathbb{N}_+ \),

\[
L^n(x, y) = w_{n-1}(y - x) = \frac{(y - x)^{n-1}}{(n-1)!}, \quad x, y \in [0, \infty), \ x \leq y
\]

(c) \( W(x, t) = e^{tx} \) for \( x \in [0, \infty) \) and \( t \in \mathbb{R} \).

**Example 2.2.** Consider the standard discrete space \((\mathbb{N}, \leq)\).

(a) For \( n \in \mathbb{N} \),

\[
w_n(x) = \binom{x+n}{n}, \quad x \in \mathbb{N}
\]

(b) For \( n \in \mathbb{N}_+ \),

\[
L^n(x, y) = w_{n-1}(y - x) = \binom{y - x + n - 1}{n - 1}, \quad x, y \in \mathbb{N}, \ x \leq y
\]

(c) \( W(x, t) = \frac{1}{(1-t)^{x+1}} \), \( x \in \mathbb{N}, \ t \in (-1, 1) \)
Exercise 2.3. Let $S = \{1, 2, 3, 4\}$ and define the symmetric relation $\leftrightarrow$ on $S$ by $1 \leftrightarrow 2, 1 \leftrightarrow 4, 2 \leftrightarrow 3, 2 \leftrightarrow 4, 3 \leftrightarrow 4$. The undirected graph $(S, \leftrightarrow)$ is known as the diamond graph and is shown in Figure 2.1. It is also an abstract version of the Wheatstone bridge graph. For the graph $(S, \leftrightarrow)$, find each of the following:

(a) The adjacency matrix $L$.
(b) The eigenvalues and corresponding eigenvectors of $L$.
(c) Find the walk function $w_n$ of order $n \in \mathbb{N}$.
(d) Find the generating function $W$.

![Figure 2.1: The diamond graph](image)

Exercise 2.4. Let $S = \{1, 2, 3, 4, 5\}$ and define the symmetric relation $\leftrightarrow$ on $S$ by $1 \leftrightarrow 2, 2 \leftrightarrow 3, 2 \leftrightarrow 4, 2 \leftrightarrow 3, 3 \leftrightarrow 5$. The undirected graph $(S, \leftrightarrow)$ is known as the bull graph and is shown in Figure 2.2. For the graph $(S, \leftrightarrow)$, find each of the following:

(a) The adjacency matrix $L$.
(b) The eigenvalues and corresponding eigenvectors of $L$.
(c) The walk function $w_n$ of order $n \in \mathbb{N}$.
(d) The generating function $W$.

![Figure 2.2: The bull graph](image)

Exercise 2.5. Let $S = \{1, 2, 3, 4\}$ and define the relation $\rightarrow$ on $S$ by $1 \rightarrow 2, a \rightarrow 3, 2 \rightarrow 4, 3 \rightarrow 4, 4 \rightarrow 1$. The graph $(S, \rightarrow)$ is a directed version of the diamond graph, and is shown in Figure 2.3. For the graph $(S, \rightarrow)$, find each of the following:

(a) The adjacency matrix $L$.
(b) Find the real eigenvalues and corresponding eigenvectors of $L$.
(c) Find the left walk function $w_n$ of order $n \in \mathbb{N}$.
(d) Find the left generating function $W$.

![Figure 2.3: The directed diamond graph](image)
2.2 Partial Orders

Of particular importance for our study are partial orders and other relations associated with partial orders. The importance stems from the fact that partial order graphs are in some ways analogous to the standard reliability setting. As usual, we have an underlying $\sigma$-finite measure space $(S, \mathcal{F}, \lambda)$ with a measurable diagonal.

**Definition 2.11.** A measurable relation $\preceq$ on $S$ that is reflexive, anti-symmetric, and transitive, is a partial order, and then the graph $(S, \preceq)$ is a partial order graph.

The term partially ordered set (poset) is commonly used in the literature, but we want to continue with graph terminology for consistency. Every partial order graph gives rise to a number of other graphs.

**Theorem 2.3.** Suppose that $(S, \preceq)$ is a partial order graph. Then the following are also graphs:

(a) $(S, \succeq)$ where $x \succeq y$ if and only if $y \preceq x$.
(b) $(S, \prec)$ where $x \prec y$ if and only if $x \preceq y$ and $x \neq y$.
(c) $(S, \succ)$ where $x \succ y$ if and only if $y \prec x$.
(d) $(S, \bot)$ where $x \bot y$ if and only if $x \preceq y$ or $y \preceq x$.
(e) $(S, \parallel)$ where $x \parallel y$ if and only if neither $x \preceq y$ nor $y \preceq x$.

**Proof.** All that is necessary is to show that the relations are measurable. As a set of ordered pairs, the equality relation $=$ is the diagonal set $\{(x,x) : x \in S\}$. By assumption, this relation is measurable. The results then follow from Propositions 2.1 and 2.2.

(a) $\succeq$ is the reverse of $\preceq$.
(b) $\prec$ is the set difference between $\preceq$ and $=$.
(c) $\succ$ is the reverse of $\prec$.
(d) $\bot$ is the union of $\preceq$ and $\succeq$.
(e) $\parallel$ is the complement of $\bot$.

Of course, $\succeq$ is also a partial order. The relations $\prec$ and $\succ$ are strict partial orders characterized by the anti-reflexive, asymmetric, and transitive properties. Also, $x \bot y$ if and only if $x$ and $y$ are comparable, and $x \parallel y$ if and only if $x$ and $y$ are non-comparable. Interval notation is natural for a partially ordered set $(S, \preceq)$. Specifically, if $a, b \in S$ then

$$[a,b] = \{x \in S : a \preceq x \preceq b\}, \quad (a,b) = \{x \in S : a < x < b\}$$
$$[a,b] = \{x \in S : a \preceq x \prec b\}, \quad (a,b) = \{x \in S : a < x \preceq b\}$$

Increasing and decreasing sets and functions make sense in partial order graphs.

**Definition 2.12.** Suppose that $(S, \preceq)$ is a partial order graph and that $A \subseteq S$. 

(a) A is increasing if $x \in A$ and $x \preceq y$ imply $y \in A$.

(b) A is decreasing if $y \in A$ and $x \preceq y$ imply $x \in A$.

**Definition 2.13.** Suppose that $(S_1, \preceq_1)$ and $(S_2, \preceq_2)$ are partial order graphs, and that $f : S_1 \to S_2$

(a) $f$ is increasing if $x, y \in S_1$ and $x \preceq_1 y$ imply $f(x) \preceq_2 f(y)$.

(b) $f$ is decreasing if $x, y \in S_1$ and $x \preceq_1 y$ imply $f(y) \preceq_2 f(x)$.

(c) $f$ is strictly increasing if $x, y \in S_1$ and $x \prec_1 y$ imply $f(x) \prec_2 f(y)$.

(d) $f$ is strictly decreasing if $x, y \in S_1$ and $x \prec_1 y$ imply $f(y) \prec_2 f(x)$.

If $(S, \preceq)$ is a partial order graph and $A \subseteq S$, then $A$ is increasing (decreasing) if and only if $1_A$ is increasing (decreasing), where $1_A$ is the indicator function of $A$ mapping $(S, \preceq)$ into $(\{0, 1\}, \leq)$. Various extremal elements make sense in a partial order graph.

**Definition 2.14.** Suppose that $(S, \preceq)$ is a partial order graph and that $A \subseteq S$.

(a) $a \in A$ is the minimum element of $A$ if $a \preceq x$ for every $x \in A$.

(b) $a \in A$ is a minimal element of $A$ if no $x \in A$ satisfies $x \prec a$.

(c) $b \in A$ is the maximum element of $A$ if $x \preceq b$ for every $x \in A$.

(d) $b \in A$ is a maximal element of $A$ if no $x \in A$ satisfies $b \prec x$.

As the terminology suggests, the minimum and maximum elements of $A$ are unique if they exist, and are denoted $\min(A)$ and $\max(A)$ respectively. On the other hand, $A$ can have multiple minimal or maximal elements (or none).

**Definition 2.15.** Suppose again that $(S, \preceq)$ is a partial order graph and that $A \subseteq S$.

(a) $u \in S$ is a lower bound of $A$ if $u \preceq x$ for all $x \in A$.

(b) $v \in S$ is an upper bound of $A$ if $x \preceq v$ for all $x \in A$.

(c) The greatest lower bound or infimum of $A$, if it exists, is the maximum of the set of lower bounds of $A$.

(d) The least upper bound or supremum of $A$, if it exists, is the minimum of the set of upper bounds of $A$.

The infimum of $A$ is unique if it exists, and is denoted $\inf(A)$. Similarly, the supremum of $A$ is unique if it exists, and is denoted $\sup(A)$. Note that every element of $S$ is a lower bound and an upper bound of $\emptyset$ since the conditions in the definition hold vacuously. For $x, y \in S$, an operator notation is commonly used for the infimum and supremum: $x \land y = \inf\{x, y\}$ and $x \lor y = \sup\{x, y\}$. When these exist, the partial order graph has a special name.

**Definition 2.16.** Suppose that $(S, \preceq)$ is a partial order graph.

(a) $(S, \preceq)$ is a lower semi-lattice if $x \land y$ exists for every $x, y \in S$.

(b) $(S, \preceq)$ is an upper semi-lattice if $x \lor y$ exists for every $x, y \in S$.

(c) $(S, \preceq)$ is a lattice if $x \land y$ and $x \lor y$ exist for every $x, y \in S$.

**Definition 2.17.** A partial order graph $(S, \preceq)$ is locally finite if $\lambda[x, y] \leq \infty$ for all $x, y \in S$.

Note that if $(S, \preceq)$ is left finite then it is locally finite since $[x, y] \subseteq \{t \in S : t \preceq y\}$ for $(x, y) \in S^2$. If $(S, \preceq)$ has a minimum element $e$ then the converse is true since $\{t \in S : t \preceq e\} \subseteq [e, x]$ for $x \in S$. Analogously, if $(S, \preceq)$ is right finite then it is locally finite, and the converse is true if $(S, \preceq)$ has a maximum element.

For the remainder of this section, we assume that $(S, \preceq)$ is a discrete, locally finite partial order graph. In this case, there is another natural relation associated with $\preceq$.

**Definition 2.18.** For $x, y \in S$, $y$ covers $x$ if $y$ is a minimal element of $\{u \in S : x \prec u\}$. If the covering relation on $S$ is denoted $\uparrow$, so that $x \uparrow y$ if and only if $y$ covers $x$, then $(S, \uparrow)$ is the covering graph or Hasse graph of $(S, \preceq)$.
For \( x \in S \) note that \( \{ y \in S : x \uparrow y \} \) is the set of elements immediately after \( x \) in the partial order while \( \{ u \in S : u \uparrow x \} \) is the set of elements immediately before \( x \) in the partial order.

**Definition 2.19.** A partial order graph \((S, \preceq)\) is uniform if for every \( x, y \in S \) with \( x \preceq y \), all paths from \( x \) to \( y \) in the covering graph \((S, \uparrow)\) have the same length. We let \( d(x, y) \) denote the common length.

**Proposition 2.7.** If \((S, \preceq)\) has minimum element \( e \), then \((S, \preceq)\) is uniform if and only if for every \( x \in S \), all paths from \( e \) to \( x \) in the covering graph \((S, \uparrow)\) have the same length.

**Proof.** If \((S, \preceq)\) is uniform, then trivially all paths from \( e \) to \( x \) in \((S, \uparrow)\) have the same length, for every \( x \in S \). Conversely, suppose that all paths from \( e \) to \( x \) have the same length in \((S, \uparrow)\) for every \( x \in S \). Suppose that \( x \preceq y \) and there are walks from \( x \) to \( y \) in \((S, \uparrow)\) with lengths \( m \) and \( n \). There must exist a path in \((S, \uparrow)\) from \( e \) to \( x \), since \((S, \preceq)\) is locally finite; let \( k \) denote the length of this path. Then we have two paths from \( e \) to \( y \) in \((S, \uparrow)\) of lengths \( k + m \) and \( k + n \), so \( k + m = k + n \) and hence \( m = n \).

**Proposition 2.8.** Suppose that \((S, \preceq)\) is uniform. For \( n \in \mathbb{N} \), let \( \uparrow^n \) denote the \( n \)-fold composition power of the covering relation \( \uparrow \) of \( \preceq \), where by convention, \( \uparrow^0 \) is the equality relation \( = \). As subsets of \( S^2 \), the partial order \( \preceq \) is the disjoint union of \( \uparrow^n \) over \( n \in \mathbb{N} \).

**Proof.** This is clear. For \( x, y \in S \), \( x \preceq y \) if and only if \( x \uparrow^n \) for some \( n \in \mathbb{N} \). Since \((S, \preceq)\) is uniform, \( x \preceq y \) if and only if \( x \uparrow^n \) for one and only one \( n \in \mathbb{N} \).

**Definition 2.20.** The dimension of a partial order graph \((S, \preceq)\) is the smallest number of chains (total orders) on \( S \) whose intersection gives \( \preceq \). We denote the dimension by \( \dim(S, \preceq) \).

### 2.3 Möbius Inversion

Suppose now that \((S, \preceq)\) is a discrete, locally finite, partial order graph. In this setting, the theory of Möbius inversion [3] is important. Let \( \mathcal{K} \) denote the collection of kernels \( K \) on \((S, \preceq)\) with \( K(x, x) \neq 0 \). So \( K \in \mathcal{K} \) means that \( K : S^2 \rightarrow \mathbb{R} \) and that \( K(x, x) \neq 0 \) for \( x \in S \) while \( K(x, y) = 0 \) if \( x \neq y \). Note that the \( I \in \mathcal{K} \), where \( I \) is the adjacency kernel of \((S, =)\), so that \( I(x, y) = 1(x = y) \) for \( (x, y) \in S^2 \). Similarly, \( L \in \mathcal{K} \) where \( L \) is the adjacency kernel of \((S, \preceq)\), so that \( L(x, y) = 1(x \preceq y) \) for \( (x, y) \in S^2 \).

**Theorem 2.4.** The space \((\mathcal{K}, \cdot)\) is a group with identity \( I \). The inverse of \( K \in \mathcal{K} \) is the function \( K^{-1} \in \mathcal{K} \) defined inductively as follows:

\[
K^{-1}(x, x) = \frac{1}{K(x, x)}, \quad x \in S
\]

\[
K^{-1}(x, y) = -\frac{1}{K(y, y)} \sum_{t \in \{x, y\}} K^{-1}(x, t)K(t, y), \quad x \prec y
\]

**Proof.** Suppose that \( J, K \in \mathcal{K} \). Since \((S, \preceq)\) is transitive, we know that \( JK \) is also a kernel for \((S, \preceq)\). In addition, \((JK)(x, x) = J(x, x)K(x, x) \neq 0 \). Hence \( JK \in \mathcal{K} \). Trivially \( I \) is the identity for \((\mathcal{K}, \cdot)\). Let \( K \in \mathcal{K} \) and let \( J \) be defined by the inverse formula in the theorem, which makes sense by the local finiteness assumption. Then

\[
(JK)(x, x) = J(x, x)K(x, x) = \frac{K(x, x)}{K(x, x)} = 1, \quad x \in S
\]

Next, if \( x \prec y \) then

\[
(JK)(x, y) = \sum_{t \in \{x, y\}} J(x, t)K(t, y) - \sum_{t \in \{x, y\}} J(x, t)K(t, y) + J(x, y)K(y, y)
\]

\[
= \sum_{t \in \{x, y\}} J(x, t)K(t, y) - \frac{1}{K(y, y)} \sum_{t \in \{x, y\}} J(x, t)K(t, y)K(y, y)
\]

\[
= \sum_{t \in \{x, y\}} J(x, t)K(t, y) - \sum_{t \in \{x, y\}} J(x, t)K(t, y) = 0
\]
So $J$ is a left inverse of $K$. Finally we show that it’s also a right inverse. First note that

$$(KJ)^2 = (KJ)(KJ) = K(JK)J = K1J = KJ$$

so that $KJ$ is idempotent. Next, since $KJ \in \mathcal{K}$, $KJ$ itself has a left inverse, say $N \in \mathcal{K}$. Then using the idempotent property,

$$KJ = I(KJ) = [N(KJ)](KJ) = N(KJ)^2 = N(KJ) = I$$

and hence $J$ is a right inverse of $K$.

Möbius inversion was first developed in the context of number theory and then extended to discrete partial order graph. In this context, the collection of kernels for $(S, \preceq)$ is known as the incidence algebra. The identity $I$ is known as the Kronecker delta function (and traditionally denoted $\delta$). The adjacency kernel $L$ of $(S, \preceq)$ is known as the Riemann function or the zeta function of the partial order graph. Like all kernels in $\mathcal{K}$, it has an inverse:

**Corollary 2.1.** The Möbius kernel $M \in \mathcal{K}$ of $(S, \preceq)$ is the inverse of the adjacency kernel $L$. It is defined inductively as follows:

$$M(x, x) = 1, \quad x \in S$$

$$M(x, y) = - \sum_{t \in[x,y)} M(x, t), \quad x \prec y$$

The theorem and corollary lead to the following fundamental result, known as the Möbius inversion formula.

**Corollary 2.2.** Suppose again that $L$ and $M$ are the adjacency kernel and Möbius kernel of $(S, \preceq)$, and that $f : S \to \mathbb{R}$.

(a) If $(S, \preceq)$ is left finite and $g = fL$ then $f = gM$.

(b) If $(S, \preceq)$ is right finite and $g = Lf$ then $f = Mg$

**Proof.** The results follow immediately since $M$ is the inverse of $L$.

(a) If $(S, \preceq)$ is left finite and $g(y) = (fL)(y) = \sum_{x \preceq y} f(x)$ for $y \in S$ then $f(y) = \sum_{x \preceq y} g(x)M(x, y)$ for $y \in S$. Note that the sums are finite.

(b) If $(S, \preceq)$ is right finite and $g(x) = (Lf)(x) = \sum_{x \preceq y} f(y)$ for $x \in S$ then $f(x) = \sum_{x \preceq y} M(x, y)f(y)$ for $x \in S$. Again, the sums are finite.

Part (a) also holds (without the left-finite assumption) if $|f|L \in \mathcal{L}_1$, that is

$$\sum_{y \in S} \sum_{x \preceq y} |f(x)| < \infty$$

Similarly, (b) holds (without the right-finite assumption) if $L|f| \in \mathcal{L}_1$, that is

$$\sum_{x \in S} \sum_{x \preceq y} |f(y)| < \infty$$

As you can see, the proof relies on the fact that the sums are well defined and can be reordered.
Chapter 3

Semigroups

Semigroups form the second basic algebraic structure that we consider in this text. Again, we start with an underlying $\sigma$-finite measure space $(S, \mathcal{S}, \lambda)$ with a measurable diagonal.

3.1 Basics

Definition 3.1. A semigroup $(S, \cdot)$ consists of the set $S$ and a measurable, associative binary operation $\cdot$ on $S$. That is,

(a) The mapping $(x, y) \mapsto xy$ from $S^2$ into $S$ is measurable.
(b) $x(yz) = (xy)z$ for all $x, y, z \in S$.

So for $x, y, z \in S$ we can write $xyz$ without ambiguity, and of course, this extends to longer finite products. Next are a couple of important definitions.

Definition 3.2. Suppose that $(S, \cdot)$ is a semigroup.

(a) An element $e \in S$ is the identity for $(S, \cdot)$ if $xe = ex = x$ for all $x \in S$.
(b) The semigroup satisfies the left cancellation law if $xy = xz$ implies $y = z$ for $x, y, z \in S$.

Definition 3.3. Suppose that $(S, \cdot)$ is a semigroup. For $x \in S$ and $A \in \mathcal{S}$ define

(a) $xA = \{xa : a \in A\}$
(b) $x^{-1}A = \{y : xy \in A\}$

In part (b), note that we are not defining $x^{-1}$ individually, but just in combination with a set $A \in \mathcal{S}$, and that $x^{-1}A$ may be empty. The following theorem is critical for various measurability requirements that we will need.

Theorem 3.1. Suppose that $(S, \cdot)$ is a semigroup and that $A \in \mathcal{S}$. Then

(a) $\{(x, y) \in S^2 : y \in xA\} \in \mathcal{S}^2$.
(b) $\{(x, y) \in S^2 : y \in x^{-1}A\} \in \mathcal{S}^2$.

Proof. Measurability and the diagonal assumption are the main tools.

(a) From the diagonal assumption and measurability it follows that $\{(x, y, xy) : x, y \in S\} \in \mathcal{S}^3$ (see [7]).

Intersecting this set with $S \times A \times S$ it follows that $\{(x, a, xa) : x \in S, a \in A\} \in \mathcal{S}^3$. Taking a cross section we then have $\{(x, xa) : x \in S, a \in A\} \in \mathcal{S}^2$, and this set is $\{(x, y) \in S^2 : y \in xA\}$.

(b) The set $\{(x, y) \in S^2 : y \in x^{-1}A\} = \{(x, y) \in S^2 : xy \in A\}$ is simply the inverse image of $A$ under the measurable mapping $(x, y) \mapsto xy$, and hence this set is measurable.

□
As a corollary, it follows that if \( x \in S \) and \( A \in \mathcal{S} \) then \( xA \in \mathcal{S} \) and \( x^{-1}A \in \mathcal{S} \). More generally, we have the following:

**Corollary 3.1.** Suppose again that \((S, \cdot)\) is a semigroup and that \( A \in \mathcal{S} \). If \( \lambda \) is a \( \sigma \)-finite measure on \((S, \mathcal{S})\) then the following functions from \( S \) to \([0, \infty)\) are measurable:

(a) \( x \mapsto \lambda(xA) \)

(b) \( x \mapsto \lambda(x^{-1}A) \)

**Proof.** These results are a direct consequence of Theorem 3.1.

(a) The function \((x, y) \mapsto 1(y \in xA)\) from \( S^2 \) to \([0, 1]\) is measurable and

\[
\lambda(xA) = \int_S 1(y \in xA) \, d\lambda(y), \quad x \in S
\]

(b) Similarly the mapping \((x, y) \mapsto 1(y \in x^{-1}A)\) from \( S^2 \) to \([0, 1]\) is measurable and

\[
\lambda(x^{-1}A) = \int_S 1(y \in x^{-1}A) \, d\lambda(y), \quad x \in S
\]

**Theorem 3.2.** Suppose that \((S, \cdot)\) is a semigroup and let \( x \in S \).

(a) The mapping \( A \mapsto x^{-1}A \) from \( \mathcal{S} \) to \( \mathcal{S} \) preserves all of the set operations.

(b) If \((S, \cdot)\) satisfies the left cancellation law, then the mapping \( A \mapsto xA \) from \( \mathcal{S} \) to \( \mathcal{S} \) preserves all of the set operations.

**Proof.** For \( x \in S \), let \( \varphi_x : S \to S \) be defined by \( \varphi_x(y) = xy \) for \( y \in S \).

(a) \( x^{-1}A = \varphi_x^{-1}(A) \) and of course inverse images of \( \varphi_x \) preserve all set operations.

(b) \( xA = \varphi_x(A) \). If \((S, \cdot)\) satisfies the left cancellation law, then \( \varphi_x \) is one-to-one and in this case, forward images of \( \varphi_x \) preserve all of the set operations.

**Theorem 3.3.** Suppose again that \((S, \cdot)\) is a semigroup and that \( x \in S \) and \( A \in \mathcal{S} \).

(a) \( A \subseteq x^{-1}(xA) \). Equality holds if \((S, \cdot)\) has the left cancellation property.

(b) \( x(x^{-1}A) = A \cap (xS) \subseteq A \).

**Proof.** As before, let \( \varphi_x : S \to S \) be defined by \( \varphi_x(y) = xy \) for \( y \in S \).

(a) By a general result for forward and inverse images of functions,

\[
A \subseteq \varphi_x^{-1}(\varphi_x(A)) = x^{-1}(xA)
\]

and equality holds if \( \varphi_x \) is one-to-one, which means that \((S, \cdot)\) has the left cancellation property.

(b) Since the co-domain of \( \varphi_x \) is \( xS \), another general result for forward and inverse images gives

\[
x(x^{-1}A) = \varphi_x(\varphi_x^{-1}(A)) = A \cap xS
\]

From part (a), it follows that if \((S, \cdot)\) satisfies the left cancellation law and if \( y \in xS \), then \( x^{-1}\{y\} \) has a single element, which we denote as \( x^{-1}y \). Note again that we are not defining \( x^{-1} \) individually, but only in combination with \( y \in xS \). The last two propositions are motivations for the following:

**Assumption 3.1.** We assume that all semigroups in this text satisfy the left cancellation law.
Corollary 3.2. Suppose that $(S, \cdot)$ is a semigroup. The mapping $(x, y) \mapsto x^{-1}y$ from $\{(x, y) : x \in S, y \in xS\}$ into $S$ is measurable.

Proof. Let $A \in \mathcal{S}$. The inverse image of $A$ under the mapping $(x, y) \mapsto x^{-1}y$ is

$$\{(x, y) : x \in S, y \in xS, x^{-1}y \in A\} = \{(x, xu) : (x, u) \in S \times A\}$$

But we know that $\{(x, u, xu) : (x, u) \in S \times A\} \in \mathcal{S}^3$ so it follows that the set in the displayed equation is in $\mathcal{S}^2$.

It follows that if $g : S \to \mathbb{R}$ is measurable, then the function $(x, y) \mapsto g(x^{-1}y)$ from $\{(x, y) : x \in S, y \in xS\}$ to $\mathbb{R}$ is measurable, a result we will use frequently.

Definition 3.4. Suppose that $(S, \cdot)$ is a semigroup and that $A \in \mathcal{S}$. The (right) relation $\to_A$ on $S$ associated with $(S, \cdot)$ and $A$ is defined by

$$x \to_A y \text{ if and only if } y = xa \text{ for some } a \in A$$

The most important case is when $A = S$. In this case, we drop the subscript and refer to $\to$ simply as the relation associated with $(S, \cdot)$. From Theorem 3.1, the relation $\to_A$ is measurable and hence $(S, \to_A)$ is a graph for every $A \in \mathcal{S}$. All of the definitions and results in the last chapter for graphs generally apply of course to the graphs associated with a semigroup. In particular, a walk of length $n \in \mathbb{N}_+$ for $(S, \to_A)$ has the form

$$(x, xa_1, xa_1a_2, \ldots, xa_1a_2 \cdots a_n) \text{ where } x \in S \text{ and } (a_1, a_2, \ldots, a_n) \in A^n$$

The following example describes a trivial semigroup, but one that is nonetheless useful for certain constructions and counterexamples.

Example 3.1. Define the binary operation $\cdot$ on $S$ by $xy = y$ for all $x, y \in S$. Then $(S, \cdot)$ is a semigroup, called the right trivial semigroup. The associated relation $\equiv$ is the complete relation on $S$: $x \equiv y$ for all $x, y \in S$.

Proof. Trivially, the mapping $(x, y) \mapsto y$ from $S^2$ to $S$ is measurable. The associative property holds: $x(yz) = xz = z$ and $(xy)z = yz = z$ for $x, y, z \in S$. The left cancellation property holds: if $xy = xz$ then by definition, $y = z$. Finally, since $xy = y$ it follows that $x \equiv y$ for all $x, y \in S$.

Theorem 3.4. Suppose again that $(S, \cdot)$ is a semigroup and let $\to_A$ denote the relation associated with $A \in \mathcal{S}$. Then $u \to_A v$ if and only if $xu \to_A xv$ for $x, u, v \in S$.

Proof. Suppose that $x, u, v \in S$ and that $u \to_A v$. Then there exists $a \in A$ such that $ua = v$. But then $x(ua) = (xu)a = xv$ so $xu \to_A xv$. Conversely, suppose that $x, u, v \in S$ and that $xu \to_A xv$. Then there exists $a \in A$ such that $(xu)a = x(ua) = xv$. By the left cancellation law, $ua = v$ so $u \to_A v$.

Corollary 3.3. Suppose that $(S, \cdot)$ is a semigroup. For $x \in S$ and $A \in \mathcal{S}$, the mapping $a \mapsto xa$ is an isomorphism from $A$ onto $xA$ with respect to the relation associated with $A$.

Proof. By definition, $a \mapsto xA$ maps $A$ onto $xA$. The mapping is one-to-one by the left cancellation property: if $xa = xb$ for $a, b \in A$ then $a = b$. Finally, by Theorem 3.4, if $u, v \in S$ then $u \to_A b$ if and only if $xu \to_A xv$.

In particular, for $x \in S$, the mapping $y \mapsto xy$ is an isomorphism from $S$ onto $xS$ with respect to the relation associated with $(S, \cdot)$.

Definition 3.5. Suppose that $(S, \cdot)$ is a semigroup and that $R \in \mathcal{S}$. Then $R$ is closed under $\cdot$ if $x, y \in R$ implies $xy \in R$. Hence $(R, \cdot)$ is a sub-semigroup of $(S, \cdot)$. If the operation is understood, we will simply say that $R$ is closed.

Theorem 3.5. Suppose that $(S, \cdot)$ is a semigroup and let $\to_R$ denote the relation on $S$ associated with a set $R \in \mathcal{S}$. Then $\to_R$ is transitive if and only if $R$ is closed.

Proof. Suppose that $R$ is closed, and that $x, y, z \in S$ with $x \to_R y$ and $y \to_R z$. Then there exists $s, t \in R$ such that $y = xs$ and $z = yt$. Hence $z = (xs)t = x(st)$. But $st \in R$ so $x \to_R z$. Conversely, suppose that $\to_R$ is transitive and that $x, y \in R$. Then $x \to_R x^2$ and $x^2 \to_R x^2y$. Hence $x \to_R x^2y$, so there exists $t \in R$ with $xt = x^2y = x(xy)$. By left cancelation, $t = xy$ so $xy \in R$. 

In particular, the relation $\to$ associated with $(S, \cdot)$ (that is, $R = S$) is transitive.

**Definition 3.6.** Suppose that $(S, \cdot)$ is a semigroup and that $\to$ is the associated relation. A sub-semigroup $(T, \cdot)$ of $(S, \cdot)$ is **complete** if $x, y \in T$ and $x \to y$ implies $x^{-1}y \in T$.

Equivalently, if $(T, \cdot)$ is complete and $\to_T$ denotes the relation associated with $(T, \cdot)$, then as sets of ordered pairs, $\to_T$ is the intersection of $\to$ with $T \times T$. Note that the definition of **complete** here is in an algebraic sense, not a topological sense.

For the following definitions (due to Székely [38]) we assume that $S$ has a locally compact, Hausdorff topology with a countable base and that $\mathcal{S}$ is the corresponding Borel $\sigma$-algebra. Of course, these assumptions automatically apply in the discrete case, corresponding to the discrete topology.

**Definition 3.7.** Suppose that $(S, \cdot)$ is a semigroup.

(a) A set $A \subseteq S$ is a **critical set** for the semigroup if the following property holds: If $\varphi$ is a continuous homomorphism from $(S, \cdot)$ into the group $(\mathbb{R}, +)$ with $\varphi(x) = 0$ for all $x \in A$ then $\varphi(x) = 0$ for all $x \in S$.

(b) The minimum cardinality of a critical set is the **dimension** of $(S, \cdot)$, denoted $\dim(S, \cdot)$.

In particular, $(S, \cdot)$ has dimension 0 if there are no non-trivial continuous homomorphisms from $(S, \cdot)$ to $(\mathbb{R}, +)$ and $(S, \cdot)$ has dimension $\infty$ if for every finite subset of $S$ there exists a nontrivial continuous homomorphism from $(S, \cdot)$ into $(\mathbb{R}, +)$ which maps the finite subset onto 0. The concept of convolution of functions makes sense in a semigroup. We are only going to need convolution for nonnegative functions.

**Definition 3.8.** Suppose that $(S, \cdot)$ is a semigroup and let $\to$ denote the relation associated with $(S, \cdot)$. For $f, g \in \mathcal{M}_+$, the **convolution** of $f$ with $g$ defined by

$$(f * g)(z) = \int_{x \to z} f(x)g(x^{-1}z) \, d\lambda(x), \quad z \in S$$

**Proof.** We need to note that the definition makes sense. Recall that $x \to z$ if and only if $z \in xS$, so $x^{-1}z$ is defined. Moreover, the function $x \mapsto g(x^{-1}z)$ from $\{x \in S : z \in xS\}$ to $[0, \infty)$ is measurable by Corollary 3.2. \qed

Like the semigroup itself, convolution is associative, but not commutative in general.

**Theorem 3.6.** Suppose again that $(S, \cdot)$ is a semigroup and let $\to$ denote the relation associated with $(S, \cdot)$. For $f, g, h \in \mathcal{M}_+$,

$$[(f * g) * h](z) = [f * (g * h)](z) = \int_{x \to y \to z} f(x)g(x^{-1}y)h(y^{-1}z) \, d\lambda^2(x,y), \quad z \in S$$

**Proof.** First,

$$[(f * g) * h](z) = \int_{y \to z} (f * g)(y)h(y^{-1}z) \, d\lambda(y) = \int_{y \to z} \left[ \int_{x \to y} f(x)g(x^{-1}y) \, d\lambda(x) \right] h(y^{-1}z) \, d\lambda(y)$$

$$= \int_{y \to z} \int_{x \to y} f(x)g(x^{-1}y)h(y^{-1}z) \, d\lambda(x) \, d\lambda(y), \quad z \in S$$

On the other hand, for $z \in S$,

$$[f * (g * h)](z) = \int_{x \to z} f(x)(g * h)(x^{-1}z) \, d\lambda(x) = \int_{x \to z} f(x) \int_{y \to x^{-1}z} g(y)h(y^{-1}(x^{-1}z)) \, d\lambda(y) \, d\lambda(x)$$

But if $x \to z$ and $y \to x^{-1}z$ then $y^{-1}(x^{-1}z) = (xy)^{-1}z$. We use the transformation $u = x, v = xy$ which transforms the domain $\{(x, y) \in S^2 : x \to z, y \to x^{-1}z\}$ one-to-one onto the domain $\{(u, v) \in S^2 : u \to v, v \to z\}$. Using the basic change of variables theorem,

$$[f * (g * h)](z) = \int_{v \to z} \int_{u \to v} f(u)g(u^{-1}v)h(v^{-1}z) \, d\lambda(u) \, d\lambda(v)$$

\qed
3.2 Positive Semigroups

Positive semigroups are a particularly important class of semigroups, since the underlying relation is a partial order.

Definition 3.9. Suppose that \((S, \cdot)\) is a semigroup.

(a) \((S, \cdot)\) is a strict positive semigroup if \(xy \neq x\) for every \(x, y \in S\).

(b) \((S, \cdot)\) is a positive semigroup if \((S, \cdot)\) has an identity element \(e\) and \((S_+, \cdot)\) is a strict positive semigroup where \(S_+ = S - \{e\}\).

Theorem 3.7. Suppose that \((S_+, \cdot)\) is a strict positive semigroup. Then \((S_+, \cdot)\) can be made into a positive semigroup with the addition of an identity element.

Proof. Note that the condition \(xy \neq x\) for all \(x, y \in S_+\) implies that \((S_+, \cdot)\) does not have an identity. Thus let \(S = S_+ \cup \{e\}\) where \(e\) is a element not in \(S\). Extend \(\cdot\) to \(S\) by \(xe = ex = x\) for all \(x \in S\). We will show that \((S, \cdot)\) is a positive semigroup. First, the associative property \((xy)z = x(yz)\) still holds, since it holds in \(S_+\) and trivially holds if one of the elements is \(e\). Next, \(e\) is the identity, by construction. Finally, \((S - \{e\}, \cdot) = (S_+, \cdot)\) is a strict positive semigroup by assumption.

In terms of the measure theory, suppose that \((S_+, \succ)\) is the underlying measure space. We define \(\mathcal{J}\) by adding to \(\mathcal{J}_+\) all sets of the form \(A \cup \{e\}\) where \(A \in \mathcal{J}_+\). The diagonal property holds:

\[
\{(x, x) : x \in S\} = \{(x, x) : x \in S_+\} \cup \{(e, e)\} \in \mathcal{J}^2
\]

Note that the algebraic assumptions of a positive semigroup do not rule out the possibility that \(xy = y\) for some \(x, y \in S\) with \(x \neq e\). The following definition is mostly of interest for discrete semigroups.

Definition 3.10. Suppose that \((S, \cdot)\) is a positive semigroup. An element \(i \in S\) is irreducible if and only if \(i\) cannot be factored \(i = xy\) for \(x, y \in S\), except for the trivial factoring \(i = ci = ie\).

Theorem 3.8. Suppose that \((S, \cdot)\) is a semigroup.

(a) If \((S, \cdot)\) is a strict positive semigroup then the associated relation \(\prec\) is a strict partial order.

(b) If \((S, \cdot)\) is a positive semigroup then the associated relation \(\preceq\) is a partial order.

Proof. The proofs are simple.

(a) Suppose that \(x \in S\) and \(x \prec x\). Then there exists \(a \in S\) with \(xa = x\), a contradiction. Hence \(\prec\) is irreflexive and since it’s transitive, also asymmetric.

(b) We first show that \(xy = e\) if and only if \(x = y = e\). Trivially if \(x = y = e\) then \(xy = e\) since \(e\) is an identity. Next, if \(x \neq e\) and \(y \neq e\) then \(xy \neq e\) since \(S_+\) is closed under \(\cdot\). Next if \(x = e\) and \(y \neq e\) then \(xy = x \neq e\). Now for the reflexive property, note that \(xe = x\) for \(x \in S\) so \(x \preceq x\). For the anti-symmetric property, suppose that \(x, y \in S\) and that \(x \preceq y\) and \(y \preceq x\). Then there exist \(a, b \in S\) such that \(xa = y\) and \(yb = x\). Hence \(x(ab) = (xa)b = yb = x\). Once again we take cases. If \(x = e\) then \(ab = e\) so \(a = b = e\) and hence \(y = e = x\). If \(x \neq e\) then by the strict positive property \(ab = e\) so once again \(a = b = e\) and hence \(x = y\).

Note that if \((S, \cdot)\) is a positive semigroup then \(e\) is the minimum element of \((S, \preceq)\) since \(e \preceq x\) for every \(x \in S\). Also, by the remark following Corollary 3.3, the mapping \(x \mapsto xy\) is an isomorphism from \(S\) to \(xS = \{y \in S : x \preceq y\}\) with respect to \(\preceq\). is The following result involves the covering relation associated with a positive semigroup, and so is mostly of interest for discrete semigroups.

Theorem 3.9. Suppose that \((S, \cdot)\) is a positive semigroup and that \(\preceq\) is the corresponding partial order. If \(x, y \in S\) then \(y\) covers \(x\) if and only if \(y = xi\) for some irreducible element \(i\). That is, the covering relation is the relation associated with \(I\), the set of irreducible elements of \((S, \cdot)\).
Proof. Suppose that \( y \) covers \( x \). Then \( x \prec y \) so there exists \( i \in S_+ \) with \( y = xi \). If \( i = ab \) for some \( a, b \in S_+ \) then \( x \prec xa \prec xab = y \) which would be a contradiction. Thus, \( i \) has no non-trivial factorings and hence is irreducible. Conversely, suppose that \( y = xi \) for some irreducible element \( i \). Then \( x \prec y \). Suppose there exists \( u \in S \) with \( x \prec u \prec y \). Then \( u = xs \) and \( y = ut \) for some \( s, t \in S_+ \). Thus \( y = xst = xi \) so by left cancellation, \( i = st \). But this is a contradiction since \( i \) is irreducible, so \( y \) covers \( x \).

\( \square \)

**Theorem 3.10.** Suppose that \((S, \cdot)\) is discrete, left-finite, positive semigroup with \( I \) as the set of irreducible elements. If \( x \in S \), then \( x \) can be factored finitely over \( I \). That is, \( x = i_1 i_2 \cdots i_n \) where \( i_k \in I \) for each \( k \).

Proof. Recall that left finite means that \( \gamma(x) < \infty \) for \( x \in S \), where as usual, \( \gamma \) is the left walk function of \((S, \leq)\). Let \( x \in S \). Since \( \gamma(x) < \infty \), there exists a finite walk in the covering graph from \( e \) to \( x \), say \((x_0, x_1, x_2, \ldots, x_n)\) where \( x_0 = e, x_n = x \), and \( x_{k+1} \) covers \( x_k \) for each \( k \). But then \( x_{k+1} = x_k i_k \) for each \( k \) where \( i_k \in I \). Hence \( x = i_1 i_2 \cdots i_n \).

Of course, the factoring of \( x \) over \( I \) is not necessarily unique, and different factorings of \( x \) over \( I \) may have different lengths. If a partial order graph \((S, \preceq)\) is associated with a positive semigroup, then \((S, \preceq)\) is uniform if and only if, for each \( x \), all factorings of \( x \) over \( I \) (the set of irreducible elements) have the same length.

**Corollary 3.4.** Suppose again that \((S, \cdot)\) is a discrete, left-finite, positive semigroup. Then \( \dim(S, \cdot) \leq \#(I) \).

Proof. Suppose that \( \varphi \) is a homomorphism from \((S, \cdot)\) into \((\mathbb{R}, +)\) and that \( \varphi(i) = 0 \) for each \( i \in I \). If \( x \in S \), then from Theorem 3.10, \( x \) can be factored over \( I \) so that \( x = i_1 i_2 \cdots i_n \) where \( i_k \in I \) for each \( k \). But then

\[
\varphi(x) = \varphi(i_1) + \varphi(i_2) + \cdots + \varphi(i_n) = 0
\]

Hence \( I \) is a critical set and so \( \dim(S, \cdot) \leq \#(I) \).

We can certainly have \( \dim(S, \cdot) < \#(I) \). The semigroup corresponding to the subset partial order on finite subsets of \( \mathbb{N}_+ \) studied in Chapter 16 has infinitely many irreducible elements but the semigroup dimension is 1. For a positive semigroup, there are two definitions for dimension, one corresponding to the semigroup itself, and one corresponding to the corresponding graph. How are they related? They are certainly not the same. Once again, the semigroup corresponding to the subset partial order on finite subsets of \( \mathbb{N}_+ \) studied in Chapter 16 provides an illustration. In this case, the semigroup dimension is 1 and partial order dimension is greater than 1.

**Problem 3.1.** Suppose that \((S, \cdot)\) is a postive semigroup with associated partial order \( \preceq \). How are \( \dim(S, \cdot) \) and \( \dim(S, \preceq) \) related?

For the remainder of this section, we assume that \((S, \cdot)\) is a discrete, left-finite positive semigroup with identity \( e \). So the theory of Möbius inversion in Section 2.3 applies to the associated partial order graph \((S, \preceq)\).

**Theorem 3.11.** Let \( M \) denote the Möbius kernel of \((S, \preceq)\). Then \( M(x, y) = M(tx, ty) \) for \( x, y, t \in S \).

Proof. The fact that \((S, \cdot)\) is left finite means that the Möbius kernel is well defined and that we can use partial order induction. Let \( x, y, t \in S \). First, \( x \preceq y \) if and only if \( tx \preceq ty \) so it follows that if \( x \not\preceq y \) then \( M(x, y) = 0 = M(tx, ty) \). Next, \( M(x, x) = 1 = M(tx, tx) \). Suppose now that \( x \prec y \). Recall that \( u \mapsto tu \) maps \([x, y]\) one to one and onto \([tx, ty]\). For the induction hypothesis, suppose that \( M(x, u) = M(tx, tu) \) for all \( u \prec y \). Then

\[
M(tx, ty) = - \sum_{z \in [tx, ty]} M(tx, z) = - \sum_{u \in [x, y]} M(tx, tu) = - \sum_{u \in [x, y]} M(x, u) = M(x, y)
\]

\( \square \)

The result is to be expected because of the self-similarity property of a positive semigroup. It follows that the general Möbius kernel \( M \) can be obtained from a simpler function.

**Definition 3.11.** The Möbius function \( \mu \) is defined by \( \mu(t) = M(e, t) \) for \( t \in S \), and is defined inductively as follows:

\[
\mu(t) = - \sum_{u \in [x, y]} M(x, u) = M(x, y)
\]
(a) \( \mu(e) = 1 \)
(b) \( \mu(x) = -\sum_{t \in [e,x]} \mu(t) \) for \( x \in S_+ \)

**Corollary 3.5.** The Möbius kernel \( M \) is related to the Möbius function \( \mu \) by
\[
M(x, y) = \mu(x^{-1}y), \quad x \leq y
\]

**Proof.** Suppose that \( x \leq y \) so that \( x^{-1}y \) makes sense. From Theorem 3.11,
\[
M(e, x^{-1}y) = M[x, x(x^{-1}y)] = M(x, y)
\]

\( \square \)

### 3.3 Left Invariance

Left invariance of the reference measure \( \lambda \) is the abstract version of translation invariance for Lebesgue measure and counting measure, and is very important for our study of probability on semigroups.

**Definition 3.12.** Suppose that \((S, \cdot)\) is a semigroup. The measure \( \lambda \) is left invariant for \((S, \cdot)\) if \( \lambda(xA) = \lambda(A) \) for every \( x \in S \) and \( A \in \mathcal{S} \).

The left cancellation property would seem to be a necessary condition for left invariance to make sense, since we have \( t \mapsto xt \) to be an isomorphism from \( S \) to \( xS \) for each \( x \in S \). We are particularly interested in the case where a left-invariant measure exists and is unique up to multiplication by positive constants.

The next two propositions give a case where uniqueness fails in the extreme, and an important case where uniqueness holds.

**Theorem 3.12.** Suppose that \((S, \cdot)\) is a right trivial semigroup. Then every measure \( \lambda \) on \((S, \mathcal{S})\) is left invariant.

**Proof.** Recall that \( xy = y \) for \( x, y \in S \). Hence \( xA = A \) for \( x \in S \) and \( A \in \mathcal{S} \) and so trivially \( \lambda(xA) = \lambda(A) \).

**Theorem 3.13.** Suppose that \((S, \cdot)\) is a discrete positive semigroup. Then \# is the unique left-invariant measure, up to multiplication by positive constants.

**Proof.** Recall that the term discrete means that \( S \) is countable and \( \mathcal{S} \) is the power set of \( S \). If \( x \in S \) and \( A \subseteq S \) then \( A \) and \( xA \) have the same cardinality so \( \#(xA) = \#(A) \). Hence \# is left invariant. On the other hand, if \( \lambda \) is a left-invariant measure then \( \lambda(\{x\}) = \lambda(x\{e\}) = \lambda(\{e\}) \) for \( x \in S \). Hence \( \lambda = \lambda(\{e\})\# \).

The following theorem and its corollaries give integral versions of left invariance.

**Theorem 3.14.** Suppose that \((S, \cdot)\) is a semigroup and that \( \lambda \) is left invariant for \((S, \cdot)\). If \( x \in S \) and \( g : \mathcal{M}_+ \) or \( g \in \mathcal{L}_1 \) then
\[
\int_{xS} g(x^{-1}y)d\lambda(y) = \int_S g(z)d\lambda(z)
\]

**Proof.** Note that \( y \mapsto g(x^{-1}y) \) is measurable, as a function from \( xS \) to \( \mathbb{R} \). If \( g = 1_A \) for \( A \in \mathcal{S} \) then
\[
\int_{xS} g(x^{-1}y)d\lambda(y) = \lambda(\{y \in xS : x^{-1}y \in A\}) = \lambda(xA), \quad x \in S
\]
since \( y \in xS \) and \( x^{-1}y \in A \) if and only if \( y \in xA \). On the other hand, of course,
\[
\int_S g(z)d\lambda(z) = \lambda(A)
\]

Since \( \lambda \) is left invariant, the condition in the theorem holds for function of the form \( g = 1_A \) for \( A \in \mathcal{S} \). By the standard bootstrapping argument, the condition holds for simple functions, for nonnegative measurable functions, and finally for integrable functions.  \( \square \)
Corollary 3.6. Suppose again that \((S, \cdot)\) is a semigroup and that \(\lambda\) is left invariant for \((S, \cdot)\). If \(x \in S\), \(A \in \mathcal{S}\), and \(h \in \mathcal{M}_+\) or \(h \in \mathcal{L}_1\) then
\[
\int_{xA} h(y) d\lambda(y) = \int_A h(xz) d\lambda(z)
\]

**Proof.** Given \(x \in S\), \(A \in \mathcal{S}\) and the function \(h\) in \(\mathcal{M}_+\) or \(\mathcal{L}_1\), define \(g : S \to \mathbb{R}\) by \(g(y) = 1_A(y) h(xy)\) for \(y \in S\). Then \(g\) is also in \(\mathcal{M}_+\) or \(\mathcal{L}_1\), respectively, and hence Theorem 3.14 applies. If \(y \in xS\) then
\[
g(x^{-1}y) = 1_A(x^{-1}y) h(y) = 1_{xA}(y) h(y)
\]
since \(x^{-1}y \in A\) if and only if \(y \in xA\). Hence
\[
\int_{xS} g(x^{-1}y) d\lambda(y) = \int_{xA} h(y) d\lambda(y)
\]

On the other hand,
\[
\int_S g(z) d\lambda(z) = \int_{1} 1_A(z) h(xz) d\lambda(z) = \int_A h(xz) d\lambda(z)
\]

Corollary 3.7. Suppose again that \((S, \cdot)\) is a semigroup and that \(\lambda\) is left invariant for \((S, \cdot)\). Suppose also that \(x \in S\) and \(A \in \mathcal{S}\) with \(A \subseteq xS\), and that \(g \in \mathcal{M}_+\) or \(g \in \mathcal{L}_1\). Then
\[
\int_A g(x^{-1}z) d\lambda(z) = \int_{x^{-1}A} g(y) d\lambda(y)
\]

**Proof.** Since \(A \subseteq xS\),
\[
\int_A g(x^{-1}z) d\lambda(z) = \int_{xS} 1_A(x) g(x^{-1}z) d\lambda(z)
\]

But for \(z \in xS\), \(z \in A\) if and only if \(x^{-1}z \in x^{-1}A\). Hence
\[
\int_{xS} 1_A(z) g(x^{-1}z) d\lambda(z) = \int_{xS} 1_{x^{-1}A}(x^{-1}z) g(x^{-1}z) d\lambda(z)
\]

Applying Theorem 3.14, the last integrals is
\[
\int_S 1_{x^{-1}A}(y) g(y) d\lambda(y) = \int_{x^{-1}A} g(y) d\lambda(y)
\]

**Proposition 3.1.** Suppose that \((S, \cdot)\) is a positive semigroup and that \(\lambda\) is left invariant for \((S, \cdot)\). Let \(e\) denote the identity, \(L\) the adjacency kernel, and \(\gamma_n\) the left walk function of order \(n \in \mathbb{N}\). Then for \(n \in \mathbb{N}_+\),
\[
L^n(x, y) = L^n(e, x^{-1}y) = \gamma_{n-1}(x^{-1}y), \quad x \preceq y
\]

**Proof.** Suppose that \(x \preceq y\) and \(n \in \mathbb{N}_+\). The simple “combinatorial argument” is that \(x \preceq x_1 \preceq x_2 \preceq \cdots \preceq x_{n-1} \preceq y\) is a walk of length \(n\) from \(x\) to \(y\) if and only if
\[
e \preceq x^{-1}x_1 \preceq x^{-1}x_2 \preceq \cdots \preceq x^{-1}x_{n-1} \preceq x^{-1}y
\]
is a walk of length \(n\) from \(e\) to \(x^{-1}y\), if and only if
\[
x^{-1}x_1 \preceq x^{-1}x_2 \preceq \cdots \preceq x^{-1}x_{n-1} \preceq x^{-1}y
\]
is a walk of length \(n - 1\) terminating in \(x^{-1}y\). As will be shown in Section 8.3, the product measure \(\lambda^{n-1}\) is left invariant for the product semigroup \((S^{n-1}, \cdot)\). Hence the three sets of walks have the same measure. \(\square\)
Example 3.2. Consider the standard continuous semigroup \(([0, \infty), +)\) with the usual Lebesgue measure structure. For \(n \in \mathbb{N}_+\),

\[
L^n(x, y) = L^n(0, y - x) = \gamma_{n-1}(y - x) = \frac{(y - x)^{n-1}}{(n-1)!}, \quad x \leq y
\]

Example 3.3. Consider the standard discrete semigroup \((\mathbb{N}, +)\). For \(n \in \mathbb{N}\),

\[
L^n(x, y) = L^n(0, y - x) = \gamma_{n-1}(y - x) = \left( y - x + \frac{n - 1}{n - 1} \right), \quad x \leq y
\]
Chapter 4

Probability on Graphs

Recall again that we have an underlying $\sigma$-finite measure space $(S, \mathcal{S}, \lambda)$ with a measurable diagonal. In this chapter we are interested in basic probability on the measure space that relates to the structure of a graph. We assume that we have a probability space $(\Omega, \mathcal{F}, P)$ in the background. All density functions are with respect to the reference measure $\lambda$.

4.1 Basics

Definition 4.1. A $\sigma$-finite measure $\mu$ on $(S, \mathcal{S})$ is (right) supported by a graph $(S, \rightarrow)$ if

$$\mu \{ y \in S : x \rightarrow y \} > 0, \hspace{1em} x \in S$$

We will often assume that measures are supported by the graph under consideration because this ensures that the measure is compatible with the graph in a fundamental way. In particular, we will usually assume that the reference measure $\lambda$ and probability distributions are supported by the graph. Suppose now that $X$ is a random variable with values in $S$.

Definition 4.2. The right probability function (RPF) $F$ of $X$ for the graph $(S, \rightarrow)$ is defined by

$$F(x) = P(x \rightarrow X), \hspace{1em} x \in S$$

Note that $F$ is measurable since the graph is measurable. If the graph supports $X$ then $F(x) > 0$ for $x \in S$. We use the right probability function, as opposed to the left, for our main definition because of the application to reliability, when the graph is a partially order space $(S, \preceq)$. In this case, the right probability function $F$ is given by the more familiar form

$$F(x) = P(X \geq x), \hspace{1em} x \in S$$

For a symmetric relation, we drop the adjective right. Of course the right probability function may not determine the distribution of $X$. Here is a trivial example:

Example 4.1. Suppose that $(S, \equiv)$ is a complete graph, so that $x \equiv y$ for every $(x, y) \in S^2$. If $X$ is a random variable with values in $S$, then the probability function $F$ of $X$ for $(S, \equiv)$ is the constant function 1:

$$F(x) = P(x \equiv X) = P(X \in S) = 1, \hspace{1em} x \in S$$

so $F$ gives no information about the distribution of $X$.

Here is our first very simple result.

Proposition 4.1. Suppose that $X$ has density function $f$ and has right probability function $F$ for $(S, \rightarrow)$. Then $F = Lf$ where $L$ is the adjacency kernel of $(S, \rightarrow)$.

Proof. Note that

$$F(x) = P(x \rightarrow X) = \int_{x \rightarrow y} f(y) d\lambda(y) = Lf(x), \hspace{1em} x \in S$$
So if $L$, operating on the right on the space $\mathcal{L}_1$, has a left inverse then the right probability function determines the distribution. However, that condition is not necessary. Here is a more complete result.

**Theorem 4.1.** Suppose again that $(S, \rightarrow)$ is a graph with adjacency kernel $L$. Then the following are equivalent:

(a) If $f$ and $g$ are probability density functions on $S$ with the same right probability function for $(S, \rightarrow)$ then $f = g$ almost everywhere, so that the distributions are the same.

(b) If $u \in \mathcal{L}_1$, $Lu = 0$ and $\int_S u \, d\lambda = 0$ then $u = 0$ almost everywhere.

**Proof.** We prove that (b) implies (a) and that the negation of (b) implies the negation of (a). So suppose first that (b) holds and that $f$ and $g$ are probability density functions with the same right probability function $F$, so that $Lf = Lg$. Then $f - g \in \mathcal{L}_1$, $L(f - g) = 0$, $\int_S (f - g) \, d\lambda = 0$. Hence $f - g = 0$ almost everywhere and so $f = g$ almost everywhere. Conversely, suppose that (b) does not hold, so that there exists $u \in \mathcal{L}_1$ with $Lu = 0$, $\int_S u \, d\lambda = 0$, and with $u$ nonzero on a set of positive measure. Then $\|u\|_1 > 0$ and hence $f = |u|/\|u\|_1$ is a probability density function on $S$. Moreover, $f \geq u/\|u\|_1$ and strict inequality most hold on a set of positive measure since $\int_S u \, d\lambda = 0$ ($u$ must be positive on a set of positive measure and negative on a set of positive measure). Hence $g := f - u/\|u\|_1$ differs from $f$ on a set of positive measure. Note that $g \geq 0$ and

$$\int_S g \, d\lambda = \int_S f \, d\lambda - \frac{1}{\|u\|_1} \int_S u \, d\lambda = 1 - 0 = 1$$

so $g$ is a probability density function. Moreover,

$$Lg = Lf - \frac{1}{\|u\|_1} Lu = Lf - 0 = Lf$$

Hence (a) does not hold. \qed

For a finite graph, $L$ is the adjacency matrix and the result can be restated very simply in terms of the spectrum of the graph.

**Corollary 4.1.** Suppose that $(S, \rightarrow)$ is a finite graph with adjacency matrix $L$. Then a right probability function on $(S, \rightarrow)$ uniquely determines the distribution except when $0$ is a (right) eigenvalue of $L$ and there is a corresponding eigenfunction that sums to $0$.

Chapter 15 on paths and grids has examples that illustrate that all of the cases can occur. Specifically, consider the undirected path on $m + 1$ vertices where $m \in \mathbb{N}_+$. If $m$ is odd, $0$ is not an eigenvalue. If $m$ is a multiple of $4$ then $0$ is a simple eigenvalue but the corresponding eigenvector has a nonzero sum. If $m = 2 \mod 4$ then again $0$ is a simple eigenvalue and the corresponding eigenvector has zero sum. Only in the last case does a right probability function not uniquely determine the distribution.

In the case of a discrete, partial order graph $(S, \preceq)$, the probability density function can sometimes be recovered from the right probability function via Möbius inversion. We defer this result until Section 4.2, since a moment condition is the natural hypothesis. The following definitions are natural.

**Definition 4.3.** Suppose that $X$ and $Y$ are random variables on $S$ with right probability functions $F$ and $G$ for $(S, \rightarrow)$, respectively. Then relative to $(S, \rightarrow)$,

(a) $X$ right dominates $Y$ if $F(x) \geq G(x)$ for $x \in S$.

(b) $X$ and $Y$ are right equivalent if $F(x) = G(x)$ for $x \in S$.

For a partial order graph $(S, \preceq)$, right dominance reduces to the usual definition of (first order) stochastic dominance: $\mathbb{P}(X \succeq x) \geq \mathbb{P}(Y \succeq x)$ for all $x \in S$. Clearly right equivalence relative to $(S, \rightarrow)$ really does define an equivalence class on the collection of probability distributions on $S$, and then right dominance can be extended to a partial order on the equivalence classes. The following is a simple result for mixtures of distributions.

**Proposition 4.2.** Suppose that $f$ and $g$ are probability density functions that have right probability functions $F$ and $G$ for $(S, \rightarrow)$ respectively. If $p \in (0, 1)$ then the density function $h = pf + (1 - p)g$ has right probability function $H = pF + (1 - p)G$ for $(S, \rightarrow)$. 
Definition 4.5. A partial order, the concepts of increasing and decreasing rate make sense.

The case where the right rate function is constant and this will be studied in Section 4.8. If the relation is rate function on its own may or may not determine the distribution of \( X \), but the right probability function along with the right rate function determine the distribution of \( X \). We are particularly interested in the case where the right rate function is constant and this will be studied in Section 4.8. If the relation is a partial order, the concepts of increasing and decreasing rate make sense.

Definition 4.4. Suppose that \( (S, \preceq) \) is a partial order graph that supports \( X \), and that \( X \) has right rate function \( r \) for \( (S, \preceq) \).

(a) \( X \) has increasing rate if \( r \) is increasing. That is, \( x \preceq y \) implies \( r(x) \leq r(y) \) for \( x, y \in S \).

(b) \( X \) has decreasing rate if \( r \) is decreasing. That is, \( x \preceq y \) implies \( r(x) \geq r(y) \) for \( x, y \in S \).

Proposition 4.3. Suppose that \( (S, \preceq) \) is a discrete, uniform positive semigroup, and let \( \uparrow \) denote the covering relation. Suppose also that \( X \) is a random variable with values in \( S \), and let \( F \) denote the right probability function of \( X \) for \( (S, \preceq) \), and for \( n \in \mathbb{N} \), let \( G_n \) the right probability function of \( X \) for \( (S, \uparrow^n) \). Then

\[
F = \sum_{n=0}^{\infty} G_n
\]

Proof. Recall that \( \uparrow^n \) is the \( n \)-fold composition power of \( \uparrow \) for \( n \in \mathbb{N} \), and in particular, \( \uparrow^0 \) is simply the equality relation =. Since \( (S, \preceq) \) is uniform, \( \preceq \) is the disjoint union of \( \uparrow^n \) over \( n \in \mathbb{N} \) (as sets of ordered pairs). Hence

\[
F(x) = \mathbb{P}(x \preceq X) = \sum_{n=0}^{\infty} \mathbb{P}(x \uparrow^n X) = \sum_{n=0}^{\infty} G_n(x), \quad x \in S
\]

Our last result of this section generalizes a theorem from Goldie and Resnick [13].

Theorem 4.2. Suppose that the graph \( (S, \rightarrow) \) is anti-symmetric and let \( S_0 = \{ x \in S : F(x) < 1 \} \). Then \( \mathbb{P}(X \in S_0) = 1 \).

Proof. Let \( C = S_0^c \) and let \( \mu \) denote the distribution of \( X \). For \( x \in C \), \( F(x) = \mu(xS) = 1 \) so \( \mu(C \cap xS) = \mu(C) \). Hence

\[
\int_C \mu(C \cap xS) \, d\mu(x) = \int_C \mu(C) \, d\mu(x) = [\mu(C)]^2
\]

Equivalently

\[
\int_C \int_C L(x, y) \, d\mu(y) \, d\mu(x) = [\mu(C)]^2
\]

where as usual, \( L \) is the adjacency kernel of the graph. By Fubini’s theorem and a change of variables,

\[
\int_C \int_C L(y, x) \, d\mu(y) \, d\mu(x) = [\mu(C)]^2
\]
Adding the last two equations we have

\[ 2\mu(C)^2 = \int_C \int_C [L(x, y) + L(y, x)] d\mu(y) d\mu(x) \]

Since \( \rightarrow \) is anti-symmetric, almost no \((x, y) \in S^2\) satisfies \(x \rightarrow y\) and \(y \rightarrow x\), so \(L(x, y) + L(y, x) \leq 1\) for almost all \((x, y) \in S^2\). Hence we have \(2\mu(C)^2 \leq \mu(C)^2\) or \(\mu(C)^2 \leq 0\). Hence \(\mu(C) = 0\).

\[ \square \]

Exercise 4.1. Consider the graph \((S, \leftrightarrow)\) defined in Exercise 2.3. Suppose that \(X\) is a random variable with values in \(S\) and with probability density function \(f\).

(a) Find the probability function \(F\) of \(X\) for \((S, \leftrightarrow)\) in terms of \(f\).

(b) Find the rate function \(r\) of \(X\) for \((S, \leftrightarrow)\) in terms of \(f\).

(c) Does \(F\) uniquely determine \(f\)?

✓✓

Exercise 4.2. Consider the graph \((S, \leftrightarrow)\) defined in Exercise 2.4. Suppose that \(X\) is a random variable with values in \(S\) and with probability density function \(f\).

(a) Find the probability function \(F\) of \(X\) for \((S, \leftrightarrow)\) in terms of \(f\).

(b) Find the rate function \(r\) of \(X\) for \((S, \leftrightarrow)\) in terms of \(f\).

(c) Does \(F\) uniquely determine \(f\)?

✓✓

Exercise 4.3. Consider the directed graph \((S, \rightarrow)\) defined in Exercise 2.5. Suppose that \(X\) is a random variable with values in \(S\) and with probability density function \(f\).

(a) Find the right probability function \(F\) of \(X\) for \((S, \rightarrow)\) in terms of \(f\).

(b) Find the right rate function \(r\) of \(X\) for \((S, \rightarrow)\) in terms of \(f\).

(c) Does \(F\) uniquely determine \(f\)?

✓✓

4.2 Moments

Suppose again that \((S, \rightarrow)\) is a graph with adjacency kernel \(L\) and that \(X\) is a random variable with values in \(S\) and right probability function \(F\) for \((S, \rightarrow)\). Natural moments of \(X\) can be found by integrating the left operator by the right probability function.

**Theorem 4.3.** If \(g \in \mathcal{M}_+\) then

\[
\int_S (gL^n)(x) F(x) d\lambda(x) = \mathbb{E}[(gL^{n+1})(X)], \quad n \in \mathbb{N}
\]

**Proof.** The main tool is simply Fubini’s theorem.

\[
\int_S (gL^n)(x) F(x) d\lambda(x) = \int_S (gL^n)(x) \mathbb{P}(x \rightarrow X) d\lambda(x)
\]

\[
= \int_S (gL^n)(x) \mathbb{E}[\mathbf{1}(x \rightarrow X)] d\lambda(x)
\]

\[
= \mathbb{E} \left[ \int_x \mathbf{1}(gL^n)(x) d\lambda(x) \right] = \mathbb{E}[(gL^{n+1})(X)]
\]

\[ \square \]
The result also holds for measurable \( g : S \to \mathbb{R} \), assuming the appropriate integrability. If \( X \) has density \( f \), the moment result in Theorem 4.3 is a simple consequence of the self-adjoint property:

\[
\langle gL^n, F \rangle = \langle gL^n, LF \rangle = \langle gL^{n+1}, f \rangle = \mathbb{E}[gL^{n+1}(X)]
\]

For the following corollary, recall that \( w_n = 1L^n \) is the left walk function of order \( n \in \mathbb{N} \) for the graph \((S, \to)\) so that \( w_0(x) = 1 \) for \( x \in S \), and for \( n \in \mathbb{N}_+ \),

\[
w_n(x) = \lambda^n \{ (x_1, x_2, \ldots, x_n) \in S^n : x_1 \to \cdots \to x_n \to x \}, \quad x \in S
\]

the measure of the set of initial parts of walks of length \( n \) that terminate in \( x \).

**Corollary 4.2.** For \( n \in \mathbb{N} \),

\[
\int_S w_n(x)F(x) \, d\lambda(x) = \mathbb{E}[w_{n+1}(X)]
\]

and in particular,

\[
\int_S F(x) \, d\lambda(x) = \mathbb{E}[w(X)] = \mathbb{E}[\lambda \{ x \in S : x \to X \}]
\]

**Proof.** This follows from Theorem 4.3 by taking \( g = 1 \). \(\square\)

**Example 4.2.** Consider the standard continuous graph \(([0, \infty), \leq)\). If \( X \) is a random variable with values in \([0, \infty)\), then Corollary 4.2 becomes

\[
\int_0^\infty \frac{x^n}{n!} F(X \geq x) \, dx = \mathbb{E}\left[\frac{X^{n+1}}{(n+1)!}\right], \quad n \in \mathbb{N}
\]

and in particular, with \( n = 1 \), the well-known result \( \mathbb{E}(X) = \int_0^\infty F(x) \, dx \). The standard continuous graph is studied in detail in Chapter 10.

**Example 4.3.** Consider the standard discrete graph \((\mathbb{N}, \leq)\). If \( X \) is a random variable with values in \( \mathbb{N} \), then Corollary 4.2 becomes

\[
\sum_{x=0}^\infty \binom{x+n}{n} F(x) = \mathbb{E}\left[\binom{X+n+1}{n+1}\right], \quad n \in \mathbb{N}
\]

and in particular, with \( n = 1 \), the well-known result \( \sum_{x=0}^\infty F(x) = \mathbb{E}(X + 1) \). The standard discrete graph is studied in detail in Chapter 11.

We now return to the problem of recovering the probability density function from the right probability function. For the following result, suppose that \((S, \leq)\) is a discrete, locally finite partial order graph. Recall that the Möbius kernel \( M \) is the inverse of the adjacency kernel \( L \). Suppose now that \( X \) is a random variable with values in \( S \), with probability density function \( f \) and with right probability function \( F \) for \((S, \leq)\). Finally, \( w \) denotes the left walk function of order 1, which in this setting is given by \( w(x) = \# \{ y \in S : y \leq x \} \).

**Theorem 4.4.** If \( \mathbb{E}[w(X)] < \infty \) then \( f = MF \). That is,

\[
f(x) = \sum_{x \geq y} M(x, y)F(y), \quad x \in S
\]

**Proof.** As noted in the remarks following Corollary 2.2, the Möbius inversion formula holds if

\[
\sum_{x \in S} Lf(x) < \infty
\]

But the sum is simply \( \mathbb{E}[w(X)] \) by Corollary 4.2. \(\square\)

Our next result involves the left generating function \( W \) of the graph \((S, \to)\). Recall that

\[
W(x, t) = \sum_{n=0}^\infty w_n(x)t^n, \quad x \in S, \ |t| < \rho(x)
\]

where \( \rho(x) \) is the radius of convergence of the power series for \( x \in S \). We can use this to define a left generating function for the random variable \( X \).
**Definition 4.6.** The left generating function $\Lambda$ of $X$ with respect to $(S, \to)$ is given by

$$\Lambda(t) = \mathbb{E}[W(X, t)] = \sum_{n=0}^{\infty} \mathbb{E}[w_n(X)] t^n, \quad |t| < \rho$$

where $\rho$ is the radius of convergence of the power series.

Note that $\Lambda(t)$ is the ordinary generating function of the sequence of moments $(\mathbb{E}[w_n(X)] : n \in \mathbb{N})$, but of course in this abstract setting, these moments do not in general determine the distribution of $X$. Nonetheless they are interesting since $\mathbb{E}[w_n(X)]$ is the expected measure of the walks of length $n$ that terminate in $X$. We can get the left generating function of $X$ by integrating the left generating function of the graph by the right probability function of $X$ for the graph:

**Theorem 4.5.** Suppose that $X$ has left generating function $\Lambda$ and right probability function $F$. Then

$$\Lambda(t) = 1 + t \int_S W(x, t)F(x) \, d\lambda(x), \quad |t| < \rho$$

**Proof.** Let $x \in S$ and $|t| < \rho$. Using Fubini’s theorem and Corollary 4.2,

$$1 + t \int_S W(x, t)F(x) \, d\lambda(x) = 1 + t \int_S \left[ \sum_{n=0}^{\infty} w_n(x) t^n \right] F(x) \, d\lambda(x)$$

$$= 1 + t \sum_{n=0}^{\infty} t^n \left[ \int_S w_n(x)F(x) \, d\lambda(x) \right] = 1 + t \sum_{n=0}^{\infty} t^n \mathbb{E}[w_{n+1}(X)]$$

$$= 1 + \sum_{n=0}^{\infty} t^{n+1} \mathbb{E}[w_n+1(X)] = \mathbb{E} \left[ \sum_{k=0}^{\infty} t^k w_k(X) \right] = \mathbb{E}[W(X, t)]$$

$\square$

**Example 4.4.** For the standard continuous graph $([0, \infty), \leq)$, $\Lambda(t) = \mathbb{E} \left( e^{tx} \right)$ so $\Lambda$ is the ordinary moment generating function of $X$. For the standard discrete graph $(\mathbb{N}, \leq)$, $\Lambda(t) = 1/(1 - t)^{X+1}$ for $|t| < 1$ so $\Lambda$ is a close cousin of the probability generating function of $X$.

The following definition and theorem concern entropy. At this point, there is no connection to the underlying relation $\rightarrow$, but there soon will be.

**Definition 4.7.** Suppose that $X$ is a random variable taking values in $S$ with density function $f$. The entropy of $X$ is

$$H(X) = -\mathbb{E} \ln[f(X)] = -\int_S f(x) \ln[f(x)] \, d\lambda(x)$$

**Theorem 4.6.** Suppose that $X$ and $Y$ are random variables on $S$ with density functions $f$ and $g$ respectively. Then

$$H(Y) = -\int_S g(x) \ln[g(x)] \, d\lambda(x) \leq -\int_S g(x) \ln[f(x)] \, d\lambda(x)$$

with equality if and only if $f(x) = g(x)$ almost everywhere (so that $X$ and $Y$ have the same distribution).

**Proof.** Note first that $\ln t \leq t - 1$ for $t > 0$, so $-\ln t \geq 1 - t$ for $t > 0$, with equality only at $t = 1$. Hence,

$$-\ln \left( \frac{f(x)}{g(x)} \right) = -\ln[f(x)] + \ln[g(x)] \geq 1 - \frac{f(x)}{g(x)}$$

for $x \in S$ with $g(x) > 0$. Multiplying by $g(x)$ gives

$$-g(x) \ln[f(x)] + g(x) \ln[g(x)] \geq g(x) - f(x), \quad x \in S$$

The inequality is trivially true if $g(x) = 0$ and hence holds on all of $S$. Therefore

$$-\int_S g(x) \ln[f(x)] \, d\lambda(x) + \int_S g(x) \ln[g(x)] \geq \int_S g(x) \, d\lambda(x) - \int_S f(x) \, d\lambda(x)$$

$$= 1 - 1 = 0$$

Equality holds if and only $f(x) = g(x)$ almost everywhere. $\square$
4.3 Random Walks

As usual, our starting point is a \( \sigma \)-finite measure space \((S, \mathcal{F}, \lambda)\) with a measurable diagonal. Density functions are with respect to \( \lambda \). Suppose now that \((S, \rightarrow)\) is a graph with adjacency kernel \( L \).

**Definition 4.8.** A (right) random walk on the graph \((S, \rightarrow)\) is a discrete-time, homogeneous Markov process \( X = (X_1, X_2, \ldots) \) with the property that, with probability 1, \( X_n \rightarrow X_{n+1} \) for all \( n \in \mathbb{N}_+ \).

Of course, the term random walk has many different meanings in different settings, and in particular, the term random walk on a graph has a different meaning in discrete graph theory. Also, it’s more common to have a discrete-time Markov process starting with index 0, but starting with index 1 is better for the purposes that we have in mind.

**Definition 4.9.** Suppose that \( X \) is a random variable supported by \((S, \rightarrow)\). The random walk on \((S, \rightarrow)\) associated with \( X \) is the random walk \( X = (X_1, X_2, \ldots) \) with the following properties:

(a) \( X_1 \) has the same distribution as \( X \).

(b) For \( n \in \mathbb{N}_+ \) and \( x \in S \), the conditional distribution of \( X_{n+1} \) given \( X_n = x \) is the same as the conditional distribution of \( X \) given \( x \rightarrow X \).

It’s clear that \( X \) really is a random walk on the graph \((S, \rightarrow)\) since \( X_n \rightarrow X_{n+1} \) for \( n \in \mathbb{N}_+ \). For the next several results, we assume that \( X \) is supported by \((S, \rightarrow)\) with density function \( f \), right probability function \( F \), and right rate function \( r \).

**Proposition 4.4.** The random walk \( X = (X_1, X_2, \ldots) \) on \((S, \rightarrow)\) associated with \( X \) has the following properties:

(a) \( X_1 \) has density \( f \).

(b) \( X \) has transition density \( P \) given by

\[
P(x, y) = \frac{f(y)}{F(x)} L(x, y), \quad (x, y) \in S^2
\]

**Proof.** Note that if \( X \) is a random variable with density \( f \), then the conditional distribution of \( X_{n+1} \) given \( X_n = x \in S \), with density \( y \rightarrow P(x, y) \), is the same as the conditional distribution of \( X \) given \( x \rightarrow X \).

So \( P(x, y) = f(y)/F(x) \) if \( x \rightarrow y \) and 0 otherwise. For the higher order transition densities, a new kernel is helpful, defined by integrating the product of the right rate function over walks. This is additional evidence the rate function as an object of general importance.

**Definition 4.10.** Let \( R_1 = L \) and for \( n \in \mathbb{N}_+ \), define \( R_{n+1} \) on \( S^2 \) by

\[
R_{n+1}(x, y) = \int_{x \rightarrow x_1 \rightarrow \cdots \rightarrow x_n \rightarrow y} r(x_1) r(x_2) \cdots r(x_n) \, d\lambda^n(x_1, x_2, \ldots, x_n), \quad (x, y) \in S^2
\]

so that the integral is over the middle parts of walks of length \( n + 1 \) from \( x \) to \( y \).

**Corollary 4.3.** Suppose again that \( X \) is the random walk on \((S, \rightarrow)\) associated with \( X \). For \( n \in \mathbb{N}_+ \), the \( n \)th order transition density \( P^n \) of \( X \) is given by

\[
P^n(x, y) = \frac{f(y)}{F(x)} R_n(x, y), \quad (x, y) \in S^2
\]

**Proof.** The result is true by definition when \( n = 1 \). For \( n \in \{2, 3, \ldots\} \),

\[
P^n(x, y) = \int_{S^{n-1}} P(x, x_1) P(x_1, x_2) \cdots P(x_{n-1}, y) \, d\lambda^{n-1}(x_1, x_2, \ldots, x_{n-1})
\]

\[
= \int_{x \rightarrow x_1 \rightarrow \cdots \rightarrow x_{n-1} \rightarrow y} \frac{f(x_1)}{F(x_1)} \frac{f(x_2)}{F(x_2)} \cdots \frac{f(x_{n-1})}{F(x_{n-1})} \frac{f(y)}{F(x_{n-1})} \, d\lambda^{n-1}(x_1, x_2, \ldots, x_{n-1})
\]

\[
= \frac{f(y)}{F(x)} \int_{x \rightarrow x_1 \rightarrow \cdots \rightarrow x_{n-1} \rightarrow y} r(x_1) r(x_2) \cdots r(x_{n-1}) \, d\lambda^{n-1}(x_1, x_2, \ldots, x_{n-1}) = \frac{f(y)}{F(x)} R^n(x, y)
\]
Corollary 4.4. If \( g \in \mathcal{M}_+ \) then \((gF)^n = f(gR_n)\) for \( n \in \mathbb{N}_+ \)

Proof. For \( y \in S \),
\[
(gF)^n(y) = \int_S g(x) F(x) P^n(x, y) d\lambda(x) = \int_S g(x) F(x) \frac{f(y)}{F(x)} R_n(x, y) d\lambda(x) \\
= f(y) \int_S g(x) R^n(x, y) d\lambda(x) = f(y)(gR_n)(y)
\]

Corollary 4.5. Suppose that \((S, \leftrightarrow)\) is a symmetric graph. Then \( fF \) is invariant for \( X \) and the corresponding normalized function is an invariant probability density function for \( X \).

Proof. Letting \( n = 1 \) and \( g = f \) in Corollary 4.4 gives \((fF)^1 = f(fL)\). But \( G := fL \) is the left probability function of \( X \), so we have \((fF)^1 = fG\). But \( G = F \) since \((S, \leftrightarrow)\) is symmetric so \( fF \), the product of the density and right probability functions, is invariant for \( X \). Moreover, \( 0 \leq fF \leq f \) so \( fF \in \mathcal{L}_1 \). Since the graph supports \( X \),
\[
\lambda\{x \in S : f(x) F(x) > 0\} = \lambda\{x \in S : f(x) > 0\} > 0
\]
so the normalized function is well defined, and is an invariant density function for \( X \).

In the discrete case, if the graph \((S, \leftrightarrow)\) is strongly connected so that the Markov chain \( X \) is irreducible, then \( X \) is positive recurrent.

Suppose for a moment that we do not have an underlying graph and that \( X \) is a discrete-time Markov process on \( S \) with transition density \( P \). Then trivially \( X \) is a random walk with respect to its own *leads-to relation* \( \rightarrow \) given by \( x \rightarrow y \) if and only if \( P(x, y) > 0 \). A natural question is when \( X \) is the random walk on the graph \((S, \rightarrow)\) associated with a probability density function \( f \). The answer is simple.

Corollary 4.6. A Markov transition density \( P \) on \( S \) has the form given in Proposition 4.4 if and only if \( P \) factors as
\[
h(x, y) = g(x) h(y), \quad x \rightarrow y
\]
for measurable functions \( g, h : S \rightarrow (0, \infty) \) with \( h \in \mathcal{L}_1 \).

Proof. Trivially the transition density \( P \) associated with probability density function \( f \) has the factoring given in (4.1), with \( g = 1/F \) and \( h = f \). Conversely, suppose that \( P \) has the factoring given in (4.1). Let \( c = \int_S h(y) d\lambda(y) \) so that \( c \in (0, \infty) \), and then define \( f \) and \( F \) by
\[
F(x) = \frac{1}{c g(x)}, \quad f(x) = \frac{h(x)}{c}, \quad x \in S
\]
Then \( f \) is a density function on \( S \) and \( P(x, y) = f(y)/F(x) \) for \( x \rightarrow y \). Moreover, since \( \int_{x \rightarrow y} P(x, y) d\lambda(y) = 1 \) for \( x \in S \), we have
\[
\int_{x \rightarrow y} f(y) d\lambda(y) = F(x), \quad x \in S
\]
so \( F \) is the right probability function of \( f \) for \((S, \rightarrow)\).

So clearly Markov processes on \( S \) that are associated with a probability density function on \( S \) are special. Nonetheless, many classical Markov chains with discrete state spaces have this form.

Consider again the original setting of a given graph \((S, \rightarrow)\) and a random variable \( X \) supported by \((S, \rightarrow)\). It’s easy to construct the random walk on \((S, \rightarrow)\) associated with \( X \). The construction is based on the standard rejection method, but first we need to deal with minor technicalities. Let \( S^* = S \cup \{\delta\} \) where \( \delta \) is a new state. Extend the relation \( \rightarrow \) to a relation \( \rightarrow_* \) on \( S^* \) by the additional rule \( x \rightarrow_* \delta \) for every \( x \in S^* \). Extend \( \mathcal{F} \) to a \( \sigma \)-algebra \( \mathcal{F}^* \) on \( S^* \) by adding the set \( \{\delta\} \cup A \) for \( A \in \mathcal{F} \).
Definition 4.11. Suppose that \( U = (U_1, U_2, \ldots) \) is a sequence of independent random variables, each with the same distribution as \( X \). Let \( N_1 = 1 \) and \( X_1 = U_1 \). Recursively, suppose that \( N_n \) and \( X_n \) for are defined for some \( n \in \mathbb{N}_+ \). Let

\[
N_{n+1} = \inf \{ k \in \mathbb{N}_+ : k > N_n \text{ and } X_n \to U_k \}
\]

where as usual, \( \inf \emptyset = \infty \). If \( N_{n+1} \in \mathbb{N}_+ \), define \( X_{n+1} = U_{N_{n+1}} \) while if \( N_{n+1} = \infty \) define \( X_{n+1} = \delta \). The sequence \( X = (X_1, X_2, \ldots) \) is the sequence of record variables associated with \( U \) and \((S, \to)\).

Theorem 4.7. The sequence of record variables \( X = (X_1, X_2, \ldots) \) is the random walk on \((S, \to)\) associated with \( X \).

Proof. By definition, \( X_1 = U_1 \) has the same distribution as \( X \). Suppose that \( n \in \mathbb{N}_+ \). Given \( X_n = x \in S \) and an event \( A \in \sigma\{X_0, \ldots, X_{n-1}\} \), the conditional distribution of \( X_{n+1} \) is the same as the distribution of \( W_N \) where \((W_1, W_2, \ldots)\) is a sequence of independent variables, each with the same distribution as \( X \), and where \( N = \min\{n \in \mathbb{N}_+ : x \to W_n\} \). In turn, the distribution of \( W_N \) is the same as the conditional distribution of \( X \) given \( x \to X \). Since \( X \) is supported by \((S, \to)\), with probability 1, \( X_n \in S \) for all \( n \in \mathbb{N} \).

Suppose again that \( X = (X_1, X_2, \ldots) \) is the random walk on \((S, \to)\) associated with \( X \), where \( X \) is supported by \((S, \to)\) and has density function \( f \), right probability function \( F \), and rate \( r \). Our interest now is in the joint distribution of \((X_1, X_2, \ldots, X_n)\) for \( n \in \mathbb{N}_+ \), and for that we need another sequence of functions, again defined by integrating the product of the right rate function over walks that terminate in a specified point.

Definition 4.12. Let \( v_0 = 1 \) and for \( n \in \mathbb{N}_+ \), define \( v_n \) on \( S \) by

\[
v_n(x) = \int_{x_1 \to \cdots \to x_n \to x} r(x_1)r(x_2) \cdots r(x_n) \, d\lambda^n(x_1, x_2, \ldots, x_n), \quad x \in S
\]

Equivalently, in operator notation, \( v_n = rv_n \) for \( n \in \mathbb{N}_+ \).

Corollary 4.7. Let \( n \in \mathbb{N}_+ \).

(a) \((X_1, X_2, \ldots, X_n)\) has density function \( g_n \) defined by

\[
g_n(x_1, x_2, \ldots, x_n) = r(x_1)r(x_2) \cdots r(x_{n-1})f(x_n), \quad x_1 \to x_2 \to \cdots \to x_n
\]

(b) \( X_n \) has density function \( f_n \) defined by

\[
f_n(x) = g_{n-1}(x)f(x), \quad x \in S
\]

(c) For \( x \in S \), the conditional density function of \((X_1, X_2, \ldots, X_n)\) given \( X_{n+1} = x \) is defined by

\[
h_n(x_1, x_2, \ldots, x_n \mid x) = \frac{1}{v_n(x)} r(x_1)r(x_2) \cdots r(x_n), \quad x_1 \to x_2 \to \cdots \to x_n \to x
\]

Proof. Since \( X_1 \) has density function \( f \) by definition, it follows that \((X_1, X_2, \ldots, X_n)\) has density function

\[
g_n(x_1, x_2, \ldots, x_n) = f(x_1)P(x_1, x_2) \cdots P(x_{n-1}, x_n) = f(x_1) \frac{f(x_2)}{F(x_1)} \cdots \frac{f(x_n)}{F(x_{n-1})} = r(x_1)r(x_2) \cdots r(x_{n-1})f(x_n), \quad x_1 \to x_2 \to \cdots \to x_n
\]

for Parts (b) and (c) follow by the usual methods.

This corollary points out once again the importance of the right rate function \( r \). In part (a), consider the special case that the set of walks of length \( n - 1 \) is a product set. Then \((X_1, X_2, \ldots, X_n)\) are independent, with \((X_1, X_2, \ldots, X_{n-1})\) identically distributed and having common density proportional to \( r \), while \( X_n \) has density \( f \).

Definition 4.13. For \( n \in \mathbb{N}_+ \), the density function \( f_n \) given in part (b) of Corollary 4.7 is the density function of order \( n \) associated with \((S, \to)\) and \( X \).
The following definition applies to random walks generally, not just random walks associated with a given distribution on $S$.

**Definition 4.14.** Let $X = (X_1, X_2, \ldots)$ be a random walk on $(S, \rightarrow)$. Define

$$N_A = \#\{n \in \mathbb{N}_+ : X_n \in A\} = \sum_{n=1}^{\infty} I(X_n \in A), \quad A \in \mathcal{F}$$

The process $N = \{N_A : A \in \mathcal{F}\}$ is the point process associated with the random walk $X$.

So $A \mapsto N_A$ is a random, discrete measure on $(S, \mathcal{F})$. Let’s return now to our standard setting of a random variable $X$ supported by $(S, \rightarrow)$ with density function $f$. Our interest is in $E(N_A)$ for $A \in \mathcal{F}$ and for that we need one more function.

**Definition 4.15.** Define $V : S \to [0, \infty]$ by

$$V(x) = \sum_{n=0}^{\infty} v_n(x), \quad x \in S$$

where $(v_n : n \in \mathbb{N})$ is the sequence of functions in Definition 4.12.

**Theorem 4.8.** Let $X$ be the random walk on $(S, \rightarrow)$ associated with $X$, and let $N$ be the corresponding point process. Then

$$E(N_A) = E[V(X); X \in A], \quad A \in \mathcal{F}$$

**Proof.** From Corollary 4.7, $P(X_n \in A) = E[v_{n-1}(X); X \in A]$ for $n \in \mathbb{N}_+$. Using Fubini’s theorem as usual,

$$E(N_A) = \sum_{n=1}^{\infty} P(X_n \in A) = \sum_{n=1}^{\infty} E[v_{n-1}(X); X \in A] = E\left[\sum_{n=0}^{\infty} v_n(X); X \in A\right] = E[V(X); X \in A]$$

So the (deterministic) measure $A \mapsto E(N_A)$ on $(S, \mathcal{F})$ has density function $Vf$.

### 4.4 Constant Rate Distributions

Recall again that our starting point is a $\sigma$-finite measure space $(S, \mathcal{F}, \lambda)$ with a measurable diagonal. Unless otherwise noted, density functions are with respect to $\lambda$. Suppose now that $(S, \rightarrow)$ is a graph and that $X$ is a random variable supported by the graph. Recall that the right probability function $F$ of $X$ for $(S, \rightarrow)$ is given by

$$F(x) = P(x \rightarrow X), \quad x \in S$$

From our support assumption, $F(x) > 0$ for $x \in S$. If $X$ has density function $f$ then

$$F(x) = \int_{x \rightarrow y} f(y) d\lambda(y), \quad x \in S$$

**Definition 4.16.** Random variable $X$ has (right) constant rate $\alpha \in (0, \infty)$ for $(S, \rightarrow)$ if the function $f = \alpha F$ is a probability density function of $X$.

So as the name suggests, the right rate function $r = f/F$ is the constant $\alpha$. Recall that the reference measure $\lambda$ is fixed in advance, and is usually a natural measure in some sense for the measurable space $(S, \mathcal{F})$. For example, counting measure is always used for discrete graphs and Lebesgue measure for Euclidean graphs. Trivially, if $X$ has constant rate $\alpha \in (0, \infty)$ with respect to $\lambda$ and if $c \in (0, \infty)$ then $X$ has constant rate $\alpha/c$ with respect to the measure $c\lambda$. More importantly, if we can choose the measure a posteriori, then every distribution has constant rate.

**Theorem 4.9.** Let $\mu$ be the measure defined by $\mu(A) = E[1/F(X), X \in A]$. Then $X$ has constant rate 1 with respect to $\mu$. 
CHAPTER 4. PROBABILITY ON GRAPHS

Proposition 4.5. Suppose that \((S, \rightarrow)\) is a discrete graph with the property that \(x \rightarrow x\) for some \(x \in S\). If there exists a distribution with constant rate \(\alpha\) then \(\alpha \in (0, 1]\).

Proof. Suppose that \(x \rightarrow x\) and that \(X\) has density function \(f\) and has right probability function \(F\) for \((S, \rightarrow)\). If \(X\) has constant rate \(\alpha \in (0, \infty)\) then

\[
  f(x) = \alpha F(x) = \alpha \left[ f(x) + \sum_{x \rightarrow y, y \neq x} f(y) \right]
\]

By the support assumption, \(F(x) > 0\) and hence \(f(x) > 0\). Rearranging the displayed equation we have

\[
  (1 - \alpha) f(x) = \alpha \sum_{x \rightarrow y, y \neq x} f(y)
\]

If \(\alpha > 1\), the left side is negative, but of course the right side is nonnegative.

In particular, this result applies to any discrete reflexive graph, and in particular, to any discrete partial order graph. The following result shows that mixtures of distributions with the same constant rate also have the common constant rate. From the point of view of linear algebra, this corresponds to an eigenvalue with multiplicity greater than one.

Proposition 4.6. Suppose that \(F\) and \(G\) are right probability functions for distributions supported by \((S, \rightarrow)\), each with constant rate \(\alpha \in (0, \infty)\) for \((S, \rightarrow)\). For \(p \in (0, 1)\), \(H = pF + (1 - p)G\) is also the right probability function for a distribution with constant rate \(\alpha\) for \((S, \rightarrow)\)

Proof. Recall that if \(f\) and \(g\) are probability density functions with corresponding right probability functions \(F\) and \(G\) for \((S, \rightarrow)\), respectively, and if \(p \in (0, 1)\), then the density function \(h = pf + (1 - p)g\) has right probability function \(H = pF + (1 - p)G\). In this case, \(f = \alpha F\) and \(g = \alpha G\) so

\[
  h = pf + (1 - p)g = \alpha [pF + (1 - p)G] = \alpha H
\]

The following result gives a simple condition for the uniform distribution to have constant rate. Recall that the graph \((S, \rightarrow)\) is right regular with parameter \(c \in (0, \infty)\) if \(\lambda\{y \in S : x \rightarrow y\} = c\) for all \(x \in S\).

Proposition 4.7. Suppose \((S, \rightarrow)\) is right regular with parameter \(c \in (0, \infty)\). If \(\lambda(S) < \infty\) then the uniform distribution on \(S\) has constant rate \(1/c\).
The complete graph is regular with parameter \( \lambda \).

Example 4.5. In particular, this applies to the complete graph \((S, \equiv)\), where \( x \equiv y \) for every \( x, y \in S \). The complete graph is regular with parameter \( \lambda(S) \), so if \( \lambda(S) < \infty \) then the uniform distribution on \( S \) has constant rate \( 1/\lambda(S) \).

Let \((S, \to)\) be a finite graph with vertices \( x, y \in S \) satisfying the following properties:

(a) \( x \to x \) and \( x \not\to z \) for \( z \in S - \{x\} \).

(b) \( y \to y \) and \( y \to w \) for some \( w \in S - \{y\} \).

Show that \((S, \to)\) does not support a constant rate distribution. \( \bigstar \)

Proposition 4.8. Suppose that \((S, \preceq)\) is a discrete, locally finite, uniform partial order graph. with the property that if \( x \to^n y \) then \( \#\{u \in S; x \to^n u, u \to^{n-1} y\} = a_n \in \mathbb{N}_+ \) for \( n \in \mathbb{N}_+ \), independent of \( x, y \in S \). If \( X \) has constant rate \( \beta \in (0, \infty) \) for \((S, \to)\) then

(a) \( X \) has constant rate \( \beta^n a_1 a_2 \cdots a_n \) for \((S, \to^n)\), for \( n \in \mathbb{N} \).

(b) \( X \) has constant rate \( \alpha \) for \((S, \preceq)\) where

\[
\frac{1}{\alpha} = \sum_{n=0}^{\infty} \frac{1}{\beta^n a_1 a_2 \cdots a_n}
\]

assuming that the infinite series is finite.

Proof. Let \( f \) denote the density function of \( X \). Let \( F \) denote the upper probability function of \( X \) for \((S, \preceq)\) and \( G_n \) the upper probability function of \( X \) for \((S, \to^n)\) for \( n \in \mathbb{N} \).

(a) We prove this by induction on \( n \). Note that \( a_1 = 1 \), so the result is true by assumption when \( n = 1 \). The result is trivial for \( n = 0 \) since \( G_0(x) = f(x) \) for \( x \in S \). Assume that the result holds for a given \( n \in \mathbb{N} \). Then

\[
G_{n+1}(x) = \sum_{x \to^{n+1} y} f(y) = \frac{1}{a_{n+1}} \sum_{x \to^n u} \sum_{u \to^1 y} f(y) = \frac{1}{a_{n+1}} \sum_{x \to^n u} G_n(u)
\]

\[
= \frac{1}{a_{n+1}} \sum_{x \to^n u} \frac{1}{\beta^n a_1 \cdots a_n} f(u) = \frac{1}{\beta^n a_1 \cdots a_{n+1}} G_1(x) = \frac{1}{\beta^{n+1} a_1 \cdots a_{n+1}} f(x), \quad x \in S
\]
(b) Recall that since \((S, \leq)\) is uniform, \(F = \sum_{n=0}^{\infty} G_n\). Hence
\[
F(x) = \sum_{n=0}^{\infty} G_n(x) = \sum_{n=0}^{\infty} \frac{1}{\beta^n a_1 \cdots a_n} f(x) = f(x) \sum_{n=0}^{\infty} \frac{1}{\beta^n a_1 \cdots a_n}, \quad x \in S
\]
\(\square\)

Note that series in Proposition 4.8 is finite if \(\beta > 1\). The basic result in Theorem 4.3 simplifies significantly when \(X\) has right constant rate. Once again, \(L\) denotes the adjacency kernel of \((S, \rightarrow)\).

**Corollary 4.8.** Suppose that \(X\) has constant rate \(\alpha \in (0, \infty)\) for \((S, \rightarrow)\) and that \(g \in \mathcal{M}_+\) or \(g \in \mathcal{L}_1\). For \(n \in \mathbb{N}\),
\[
\mathbb{E}(gL^n)(X) = \frac{1}{\alpha^n} \mathbb{E}(g(X))
\]

**Proof.** By assumption, \(f = \alpha F\) is a density of \(X\), so
\[
\int_S (gL^n)(x)F(x) \, d\lambda(x) = \frac{1}{\alpha} \int_S (gL^n)(x)f(x) = \frac{1}{\alpha} \mathbb{E}(gL^n)(X)
\]
Hence from Theorem 4.3
\[
\frac{1}{\alpha} \mathbb{E}(gL^n)(X) = \mathbb{E}((gL^{n+1})(X)]
\]
The result then follows since \(L^0 = I\), the identity kernel. \(\square\)

Recall that the left walk function \(w_n\) of order \(n \in \mathbb{N}\) is given by
\[
w_n(x) = 1L^n(x) = \lambda^n \{ (x_1, x_2, \ldots, x_n) \in S^n : x_1 \rightarrow x_2 \rightarrow \cdots \rightarrow x_n \rightarrow x, \quad x \in S
\]
the measure of the initial parts of the walks of length \(n\) that terminate in \(x\).

**Corollary 4.9.** If \(X\) has constant rate \(\alpha \in (0, \infty)\) then
\[
\mathbb{E}(w_n(X)) = \frac{1}{\alpha^n}, \quad n \in \mathbb{N}
\]

**Proof.** This follows from Corollary 4.8 with \(g = 1\). \(\square\)

From Corollary 4.9 note that
\[
\int_S \alpha^n w_n(x)f(x) \, d\lambda(x) = 1
\]
so it follows that the function \(f_n\) given by \(f_n(x) = \alpha^n w^n(x)f(x)\) for \(x \in S\) is a probability density function. We will return to this point shortly.

**Corollary 4.10.** Suppose that \(X\) has constant rate \(\alpha > 0\). Then
(a) \(w_n(x) < \infty\) for almost all \(x \in S\).
(b) If \(w(x) = c \in (0, \infty)\) for almost all \(x \in S\) then \(\alpha = 1/c\).

**Proof.** Recall that the graph \((S, \rightarrow)\) supports the distribution of \(X\).
(a) If \(\lambda \{ x \in S : w(x) = \infty \} > 0\) then \(\mathbb{P}[w(X) = \infty] > 0\) and hence \(\mathbb{E}[w(X)] = \infty\). This contradicts Corollary 4.9.
(b) If \(w(x) = c\) for almost all \(x \in S\) then \(\mathbb{P}[w(X) = c] = 1\) so \(\mathbb{E}[w(X)] = \mathbb{E}[w(X); w(X) = c] = c\). Hence \(c = 1/\alpha\). \(\square\)

Note that the assumption in part (b) means that \((S, \rightarrow)\) is left \(c\)-regular. The generating function result also simplifies when the distribution has constant rate.
Theorem 4.11. Suppose that $X$ has right constant rate $\alpha \in (0, \infty)$ for $(S, \rightarrow)$. Then the left generating function $\Lambda$ of $X$ is given by

$$\Lambda(t) = \frac{\alpha}{\alpha - t}, \quad |t| < \alpha$$

Proof. Recall that $\Lambda(t) = \mathbb{E}[W(X, t)]$ where $W(x, t) = \sum_{n=0}^{\infty} w_n(x)t^n$ is the left generating function of $(S, \rightarrow)$. Hence using Corollary 4.9 and Fubini’s theorem we have

$$\Lambda(t) = \mathbb{E}[W(X, t)] = \sum_{n=0}^{\infty} \mathbb{E}[w_n(X)] t^n = \sum_{n=0}^{\infty} (t/\alpha)^n = \frac{1}{1 - t/\alpha}, \quad |t| < \alpha$$

Constant rate distributions maximize entropy among distributions satisfying a certain moment condition. This is an indication that a constant rate distribution governs the most random way to put points in a graph, but we will see a stronger expression of that idea shortly.

Theorem 4.12. Suppose that $X$ has constant rate and right probability function $F$ for the graph $(S, \rightarrow)$. Then $X$ maximizes entropy among all random variables $Y$ on $S$ satisfying

$$\mathbb{E}[\ln F(Y)] = \mathbb{E}[\ln F(X)]$$

Proof. Suppose that $X$ has constant rate $\alpha$ with respect to $\lambda$, so that $f = \alpha F$ is a density of $X$. Suppose that $Y$ is a random variable taking values in $S$, with density function $g$. From the general entropy inequality in Proposition 4.6 we have

$$H(Y) = -\int_S g(x) \ln[g(x)] d\lambda(x) \leq -\int_S g(x) \ln[f(x)] d\lambda(x)$$

$$= -\int_S g(x)(\ln(\alpha) + \ln[F(x)]) d\lambda(x)$$

$$= -\ln(\alpha) - \int_S g(x) \ln[F(x)] d\lambda(x) = -\ln(\alpha) - \mathbb{E}(\ln[F(Y)])$$

Of course, the entropy of $X$ achieves the upper bound.

Note that since the right probability function $F$ typically has an “exponential” form of some sort, $\mathbb{E}(\ln[F(Y)])$ often reduces to a natural moment condition.

Results related to the random walk also simplify significantly when the underlying distribution has constant rate. For the next two results, suppose that $X$ has constant rate $\alpha \in (0, \infty)$ for $(S, \rightarrow)$ and has right probability function $F$. Suppose that $X = (X_1, X_2, \ldots)$ is the random walk on $(S, \rightarrow)$ associated with $X$.

Corollary 4.11. The transition density $P_n$ of order $n \in \mathbb{N}_+$ for $X$ is given by

$$P^n(x, y) = \alpha^n \frac{F(y)}{F(x)} L^n(x, y), \quad (x, y) \in S^2$$

If the graph $(S, \rightarrow)$ is symmetric, $F^2$ is invariant for $X$.

Proof. Recall from Corollary 4.3 that

$$P^n(x, y) = \frac{f(y)}{F(x)} R_n(x, y), \quad (x, y) \in S^2$$

where

$$R_n(x, y) = \int_{x \rightarrow x_1 \rightarrow \cdots \rightarrow x_{n-1} \rightarrow y} r(x_1) \cdots r(x_{n-1}) d\lambda(x_1, \ldots, x_{n-1}), \quad (x, y) \in S^2$$

Since $X$ has constant rate $\alpha$, $f(x) = \alpha F(x)$ and $r(x) = \alpha$ for $x \in S$. Hence

$$R_n(x, y) = \alpha^{n-1} \lambda^{n-1} \{x_1, \ldots, x_{n-1} : x \rightarrow x_1 \rightarrow \cdots \rightarrow x_{n-1} \rightarrow y\} = \alpha^{n-1} L^n(x, y)$$
Corollary 4.12. Let $n \in \mathbb{N}_+.$

(a) $(X_1, X_2, \ldots, X_n)$ has density function $g_n$ given by

$$g_n(x_1, x_2, \ldots, x_n) = \alpha^n F(x_n), \ x_1 \to x_2 \to \cdots \to x_n$$

(b) $X_n$ has density function $f_n$ given by

$$f_n(x) = \alpha^n w_{n-1}(x) F(x), \ x \in S$$

(c) For $x \in S$, the conditional distribution of $(X_1, X_2, \ldots, X_n)$ given $X_{n+1} = x$ is uniform on

$$\{(x_1, x_2, \ldots, x_n) \in S^n : x_1 \to x_2 \to \cdots \to x_n \to x\}$$

Proof. The results follow easily from the constant rate property.

(a) This follows since $r(x) = \alpha$ and $f(x) = \alpha F(x)$ for $x \in S$.

(b) This follows since $v_n(x) = \alpha^n w_n(x)$ and again $f(x) = \alpha F(x)$ for $x \in S$.

(c) The conditional distribution reduces to

$$h(x_1, x_2, \ldots x_n \mid x) = \frac{1}{w_n(x)}, \ x_1 \to x_2 \to \cdots \to x_n \to x$$

For $n \in \mathbb{N}_+$, the distribution of $X_n$ depends only on the rate parameter $\alpha$, the left left walk function $w_{n-1}$ and the right right probability function $F$. Although a simple corollary, part (c) is one of our fundamental results, so we will restate it for emphasis:

Theorem 4.13. The random walk $X$ associated with a constant rate distribution (if such a distribution exists) is the most random way to put points in the graph $(S, \to)$.

In the special case of a finite, strongly connected graph $(S, \to)$, it follows from Theorem 4.10 that the normalized eigenfunction corresponding to the largest eigenvalue of the adjacency matrix $L$ defines a distribution whose random walk is the most random way to put points in $(S, \to)$.

The next result refers to the point process $N = \{N_A : A \in \mathcal{A}\}$ corresponding to the random walk $X$. The expected number of points in a set has a simple expression in terms of the left generating function.

Corollary 4.13. Suppose that $(S, \to)$ has left generating function $W$ and that $X$ has right constant rate $\alpha \in (0, \infty)$ for $(S, \to)$. Let $X$ denote the random walk on $(S, \to)$ associated with $X$. For $A \in \mathcal{A}$,

$$\mathbb{E}(N_A) = \mathbb{E}[W(X, \alpha); X \in A]$$

Proof. This follows from Theorem 4.8 since

$$V(x) = \sum_{n=0}^{\infty} v_n(x) = \sum_{n=0}^{\infty} \alpha^n w_n(x) = W(x, \alpha), \ x \in S$$

Our next topic is thinning the point process. As above, suppose that $X = (X_1, X_2, \ldots)$ is the random walk on $(S, \to)$ corresponding to a distribution with constant rate $\alpha \in (0, \infty)$ and right probability function $F$ (and hence density $f = \alpha F$). Let $N$ have the geometric distribution on $\mathbb{N}_+$ with success parameter $p \in (0, 1)$ so that

$$\mathbb{P}(N = n) = p(1 - p)^{n-1}, \ n \in \mathbb{N}_+$$

As we will see in Chapter 11, $N$ has constant rate for the graph $(\mathbb{N}_+, \leq)$. Moreover, we assume that $N$ and $X$ are independent. The basic idea is that we accept a point with probability $p$ and reject the point with probability $1 - p$, so that $X_N$ is the first point accepted. We are interested in the distribution of $X_N$. As before, let $W$ denote the generating function of $(S, \to)$. 
CHAPTER 4. PROBABILITY ON GRAPHS

Theorem 4.14. The density function \( g \) of \( X_N \) is is given by

\[
g(x) = pW[x, \alpha(1-p)]f(x), \quad x \in S
\]

Proof. Let \( w_n \) denote the left walk function of \((S, \to)\) of order \( n \in \mathbb{N} \). Then

\[
g(x) = \mathbb{E}[f_N(x)] = \sum_{n=1}^{\infty} p(1-p)^{n-1} f_n(x) = \sum_{n=1}^{\infty} p(1-p)^{n-1} \alpha^n w_{n-1}(x) F(x)
\]

\[
= \alpha p F(x) \sum_{n=1}^{\infty} [\alpha(1-p)]^{n-1} w_{n-1}(x) = \alpha p F(x) W[x, \alpha(1-p)]
\]

\[
= pW[x, \alpha(1-p)]f(x), \quad x \in S
\]

Example 4.6. For the standard continuous graph \((0, \infty), \leq\) recall that the left walk function \( w_n \) of order \( n \in \mathbb{N} \) is given by \( w_n(x) = x^n/n! \) for \( x \in [0, \infty) \) and the generating function \( W \) is given by \( W(x, t) = e^{xt} \) for \( x \in [0, \infty) \) and \( t \in \mathbb{R} \).

(a) Random variable \( X \) has constant rate \( \alpha \in (0, \infty) \) if and only if \( X \) has the exponential distribution, with density function \( f \) and right probability function \( F \) given by \( f(x) = \alpha e^{-\alpha x} \) and \( F(x) = e^{-\alpha x} \) for \( x \in [0, \infty) \).

(b) The random walk \( \mathbf{X} = (X_1, X_2, \ldots) \) associated with \( X \) has transition density \( P^n \) of order \( n \in \mathbb{N} \) given by

\[
P^n(x, y) = \alpha^n \frac{(y-x)^{n-1}}{(n-1)!} e^{-\alpha(y-x)}, \quad x \leq y
\]

(c) For \( n \in \mathbb{N}_+ \), the density function \( g_n \) of \((X_1, X_2, \ldots, X_n)\) is given by

\[
g_n(x_1, x_2, \ldots, x_n) = \alpha^n e^{-\alpha x_n}, \quad 0 \leq x_1 \leq \cdots \leq x_n
\]

(d) For \( n \in \mathbb{N}_+ \), the density function of \( X_n \) is given by

\[
f_n(x) = \alpha^n \frac{x^{n-1}}{(n-1)!} e^{-\alpha x}, \quad x \in [0, \infty)
\]

(e) \( \mathbb{E}(N_A) = \alpha \lambda(A) \) for measurable \( A \subseteq [0, \infty) \).

(f) With thinning parameter \( p \in (0, 1) \), \( X_N \) has density function \( g \) given by

\[
g(x) = pW[x, \alpha(1-p)]f(x) = \alpha p e^{-\alpha xp}, \quad x \in [0, \infty)
\]

Of course, all of this is well known. The random walk \( \mathbf{X} \) is the sequence of arrival times of the Poisson process with rate \( \alpha \).

Example 4.7. For the standard discrete graph \((\mathbb{N}, \leq)\) recall that the left walk function \( w_n \) of order \( n \in \mathbb{N} \) is given by \( w_n(x) = \binom{x+n}{n} \) for \( x \in \mathbb{N} \) and the generating function \( W \) is given by \( W(x, t) = 1/(1-t)^{x+1} \) for \( x \in \mathbb{N} \) and \( t \in (-1, 1) \).

(a) Random variable \( X \) has constant rate \( \alpha \in (0, 1) \) if and only if \( X \) has the geometric distribution, with density function \( f \) and right probability function \( F \) given by \( f(x) = \alpha(1-\alpha)^x \) and \( F(x) = (1-\alpha)^x \) for \( x \in \mathbb{N} \).

(b) The random walk \( \mathbf{X} = (X_1, X_2, \ldots) \) associated with \( X \) has transition density \( P^n \) of order \( n \in \mathbb{N} \) given by

\[
P^n(x, y) = \alpha^n \frac{(y-x)^{n-1}}{(n-1)!} e^{-\alpha(y-x)}, \quad x \leq y
\]
(c) For \( n \in \mathbb{N}_+ \), the density function \( g_n \) of \((X_1, X_2, \ldots, X_n)\) is given by
\[
 g_n(x_1, x_2, \ldots, x_n) = \alpha^n (1 - \alpha)^{x_n}, \quad 0 \leq x_1 \leq \cdots \leq x_n
\]

(d) For \( n \in \mathbb{N}_+ \), the density function of \( X_n \) is given by
\[
 f_n(x) = \alpha^n \left( \frac{x + n - 1}{n - 1} \right) (1 - \alpha)^x, \quad x \in \mathbb{N}
\]

(e) \( E(N_A) = \frac{\alpha}{1 - \alpha} \#(A) \) for measurable \( A \subseteq \mathbb{N} \).

(f) With thinning parameter \( p \in (0, 1) \), \( X_N \) has density function \( g \) given by
\[
 g(x) = pW[x, \alpha(1 - p)] f(x) = \frac{rp}{1 - r + rp} \left( \frac{1 - r}{1 - r + rp} \right)^x, \quad x \in \mathbb{N}
\]

Of course, all of this is well known. For \( n \in \mathbb{N}_+ \), \( X_n \) is the number of failures before the \( n \) success in a Bernoulli trials sequence with success parameter \( \alpha \), so \( X \) can be thought of as the sequence of arrival times in a renewal process.

However, in general, \( X_N \) does not have constant rate.

**Exercise 4.5.** Recall the symmetric graph \((S, \leftrightarrow)\) studied in Exercise 2.3.

(a) Find the rate constant of the random variable \( X \) that has constant rate for \((S, \leftrightarrow)\).

(b) Find the density function \( f \) and probability function \( F \) of \( X \).

(c) Find the density function \( f_n \) of \( X_n \) for \( n \in \mathbb{N}_+ \), where \( X = (X_1, X_2, \ldots) \) is the random walk on \((S, \leftrightarrow)\) corresponding to \( X \).

(d) Find \( \lim_{n \to \infty} f_n \) and verify that the limit is the invariant density function of \( X \).

**Exercise 4.6.** Recall the symmetric graph \((S, \leftrightarrow)\) studied in Exercise 2.4.

(a) Find the rate constant of the random variable \( X \) that has constant rate for \((S, \leftrightarrow)\).

(b) Find the density function \( f \) and probability function \( F \) of \( X \).

(c) Find the density function \( f_n \) of \( X_n \) for \( n \in \mathbb{N}_+ \), where \( X = (X_1, X_2, \ldots) \) is the random walk on \((S, \leftrightarrow)\) corresponding to \( X \).

(d) Find \( \lim_{n \to \infty} f_n \) and verify that the limit is the invariant density function of \( X \).

**Exercise 4.7.** Recall the directed graph \((S, \rightarrow)\) studied in Exercise 2.5 and Exercise 4.3.

(a) Find the rate constant of the random variable \( X \) that has constant rate for \((S, \rightarrow)\).

(b) Find the density function and right probability function of \( X \).

(c) Find the density function \( f_n \) of \( X_n \) for \( n \in \mathbb{N}_+ \), where \( X = (X_1, X_2, \ldots) \) is the random walk on \((S, \rightarrow)\) corresponding to \( X \).

(d) Classify the random walk \( X \) in terms of periodicity, and explain the limiting behavior of \( X \).
Chapter 5

Probability on Semigroups

5.1 Basics

Our starting point in this chapter is a measurable space \((S, \mathcal{S})\) with a measurable diagonal, and a measurable semigroup \((S, \cdot)\). Recall that the relation \(\rightarrow_A\) associated with \((S, \cdot)\) and a set \(A \in \mathcal{S}\) is defined by \(x \rightarrow_A y\) if and only if \(y \in xA\). We are mostly interested in the case where \(A = S\), in which case we drop the reference to \(A\), so that \((S, \rightarrow)\) is the graph associated with the semigroup \((S, \cdot)\).

**Definition 5.1.** A \(\sigma\)-finite measure \(\mu\) on \((S, \mathcal{S})\) is supported by \((S, \cdot)\) if \(\mu\) is supported by the associated graph \((S, \rightarrow)\), so that \(\mu(xS) > 0\), \(x \in S\).

We will usually assume that measures, and in particular probability distributions on \(S\) are supported by \((S, \cdot)\) so that the measures are connected to the semigroup in a fundamental way. All of the results in the previous chapter apply to the graph \((S, \rightarrow)\) so in particular, the right probability function \(F\) of random variable \(X\) is given by

\[
F(x) = P(x \rightarrow X) = P(X \in xS), \quad x \in S
\]

But there are other probability concepts that stem from the semigroup operation more directly. For the remainder of this section, we assume that \(\lambda\) is a fixed \(\sigma\)-finite reference measure on \((S, \mathcal{S})\) and that \(\lambda\) is left \(\lambda\)-invariant for \((S, \cdot)\). All densities are with respect to \(\lambda\). The following simple proposition and its corollaries concern the distribution of products.

**Proposition 5.1.** Suppose that \(Y\) is a random variables with values in \(S\) and with density function \(g\). For \(x \in S\), a density of \(xY\) is given by

\[
z \mapsto g(x^{-1}z) \text{ for } y \in xS.
\]

**Proof.** Clearly \(xY\) takes values in \(xS\). Let \(A \in \mathcal{S}\) with \(A \subseteq xS\). Then by the integral versions of left invariance in Corollary 3.7,

\[
P(xY \in A) = P(Y \in x^{-1}A) = \int_{x^{-1}A} g(y)d\lambda(y) = \int_A g(x^{-1}z)d\lambda(z)
\]

**Corollary 5.1.** Suppose that \(X\) and \(Y\) are independent random variable with values in \(S\) and with density functions \(f\) and \(g\) respectivley. Then \((X, XY)\) has density function \(h\) given by

\[
h(x, z) = f(x)g(x^{-1}z), \quad x \rightarrow z
\]

**Proof.** Clearly \((X, XY)\) takes values in \(\{(x, xy) : (x, y) \in S^2\} = \{(x, z) \in S^2 : x \rightarrow z\}\). Moreover \(h(x, z)\) can be expressed as \(f(x)\) times the conditional density of \(XY\) at \(z\) given \(X = x\). But by independence, this is just the density of \(xY\) at \(z\). So the result follows from Proposition 5.1.

**Corollary 5.2.** Suppose again that \(X\) and \(Y\) are independent random variables with values in \(S\) and with density functions \(f\) and \(g\) respectively. Then \(XY\) has density function \(f \ast g\), the convolution of \(f\) with \(g\):

\[
(f \ast g)(z) = \int_{x \rightarrow z} f(x)g(x^{-1}z)d\lambda(z), \quad z \in S
\]
Proof. This follows immediately from Corollary 5.1. \qed

**Corollary 5.3.** Suppose that \((X_1, X_2, \ldots)\) is a sequence of independent random variables with values in \(S\) and that \(X_i\) has density function \(f_i\) for \(i \in \mathbb{N}_+\). For \(n \in \mathbb{N}_+\) let \(Y_n = X_1 \cdots X_n\). Then \((Y_1, Y_2, \ldots, Y_n)\) has density function \(h_n\) given by

\[
h_n(y_1, y_2, \ldots, y_n) = f_1(y_1)f_2(y_1^{-1}y_2) \cdots f_n(y_{n-1}^{-1}y_n), \quad y_1 \to y_2 \to \cdots \to y_n
\]

Proof. This follows by repeated application of the first corollary 5.1. \qed

The concept of compound Poisson distribution makes sense in the semigroup setting.

**Definition 5.2.** Suppose that \(X\) is a random variable with values in \(S\). Then \(X\) has a **compound Poisson distribution** if \(X\) can be factored as

\[
X = V_1 V_2 \cdots V_N
\]

where \(V = (V_1, V_2, \ldots)\) is a sequence of independent and identically distributed variables with values in \(S\), and where \(N\) is independent of \(V\) and has a Poisson distribution. The common distribution of the sequence \(V\) is the **compound distribution**.

Suppose now that \((T, \cdot)\) is a (measurable) sub-semigroup of \((S, \cdot)\). Suppose also the \(X\) is a random variable with values in \(S\) and with \(\mathbb{P}(X \in T) > 0\). We collect some simple facts about the conditional distribution of \(X\) given \(X \in T\). Recall that the right probability function \(F\) of \(X\) for \((S, \cdot)\) is defined by

\[
F(x) = \mathbb{P}(X \in xS), \quad x \in S
\]

Given \(X \in T\), the right probability function \(F_T\) of \(X\) for \((T, \cdot)\) is defined by

\[
F_T(x) = \mathbb{P}(X \in xT \mid X \in T) = \frac{\mathbb{P}(X \in xT)}{\mathbb{P}(X \in T)}, \quad x \in T
\]

If \(X\) has density function \(f\) then the density \(f_T\) of \(X\) given \(X \in T\) is defined by

\[
f_T(x) = \frac{f(x)}{\mathbb{P}(X \in T)}, \quad x \in T
\]

Consider the special case where \((S, \cdot)\) is a discrete positive semigroup, with identity element \(e\). As usual, let \(S_+ = \{x \in S : x \neq e\}\) so that \((S_+, \cdot)\) is a strict positive semigroup. If \(f\) is the density of \(X\) then the density \(f_+\) of \(X\) given \(X \in S_+\) is defined by

\[
f_+(x) = \frac{f(x)}{1 - f(e)}, \quad x \in S_+
\]

If \(F\) is the right probability function of \(X\) for \((S, \cdot)\) then the right probability function of \(X\) given \(X \in S_+\) for \((S_+, \cdot)\) is defined by

\[
F_+(x) = \frac{F(x) - f(e)}{1 - f(e)}, \quad x \in S_+
\]

We are restricting our attention to discrete positive semigroups because in the continuous case typically \(\mathbb{P}(X = e) = 0\) so the conditional distributions are the same as the unconditional distributions.

### 5.2 Random Walks

Suppose that \(X\) is a random variable supported by \((S, \cdot)\).

**Definition 5.3.** The **random walk** on \((S, \cdot)\) associated with \(X\) is the discrete time, homogeneous Markov process \(X = (X_1, X_2, \ldots)\) on \(S\) satisfying the following properties:

(a) \(X_1\) has the same distribution as \(X\).

(b) For \(n \in \mathbb{N}_+\) and \(x \in S\), the conditional distribution of \(X_{n+1}\) given \(X_n = x\) is the same as the distribution of \(xX\).
Note that $\mathbf{X}$ is also a random walk on $(S, \rightarrow)$ since $X_{n+1} \in X_n S$ or equivalently $X_n \rightarrow X_{n+1}$ for $n \in \mathbb{N}_+$. It’s trivial to construct the random walk on $(S, \cdot)$ associated with $X$.

**Proposition 5.2.** Let $\mathbf{U} = (U_1, U_2, \ldots)$ be a sequence of independent variables with values in $S$, each with the distribution of $X$. Define $X_n = U_1 \cdots U_n$ for $n \in \mathbb{N}_+$, so that $\mathbf{X}$ is the partial product sequence associated with $\mathbf{U}$. Then $\mathbf{X}$ is the random walk on $(S, \cdot)$ associated with $X$.

**Proof.** For $n \in \mathbb{N}_+$ note that $X_{n+1} = X_n U_{n+1}$ and that $U_{n+1}$ is independent of $(U_1, \ldots, U_n)$ and hence also $(X_1, \ldots, X_n)$. So it’s clear that $\mathbf{X}$ is a discrete-time Markov process. By definition, $X_1 = U_1$ has the same distribution as $X$. Moreover, the conditional distribution of $X_{n+1}$ given $X_n = x \in S$ is the same as the distribution of $xU_{n+1}$, which is the same as the distribution of $xX$.

**Proposition 5.3.** Suppose that $X$ has density function $f$. The random walk $\mathbf{X} = (X_1, X_2, \ldots)$ on $(S, \cdot)$ associated with $X$ has transition density $Q$ given by

$$Q(x, y) = f(x^{-1}y), \quad x \in S, \ y \in xS$$

**Proof.** For $x \in S$, the conditional density of $X_{n+1}$ given $X_n = x$ is the same as the density of $xX$, which by Theorem 5.1 is $y \mapsto f(x^{-1}y)$ on $xS$.

Suppose again that $X$ has density function $f$ and has right probability function $F$ for $(S, \rightarrow)$. We now have two random walks associated with $X$: one on the graph $(S, \rightarrow)$ with transition density $P$ given by $P(x, y) = f(y)/F(x)$ for $x \in S$ and $y \in xS$ and the other on the semigroup $(S, \cdot)$ itself with transition density $Q$ given by $Q(x, y) = f(x^{-1}y)$ again for $x \in S$ and $y \in xS$. The first can be constructed from an independent, identically distributed sequence $\mathbf{U} = (U_1, U_2, \ldots)$ via record variables while the second can be constructed from the sequence $\mathbf{U}$ via partial products. When are these two random walks the same? The answer is given in the next section.

**Exercise 5.1.** Consider the standard continuous semigroup $((0, \infty), +)$ with $\leq$ as the associated relation (and with Lebesgue measure as the invariant measure). Suppose that $X$ has density function $f$ given by

$$f(x) = \frac{a}{(x + 1)^{a+1}}, \quad x \in [0, \infty)$$

where $a \in (0, \infty)$ is a parameter. So $X$ has a version of the Pareto distribution with parameter $a$.

(a) Find the right probability function $F$ of $X$ for $([0, \infty), \leq)$.

(b) Find the transition density $P$ of the random walk on $([0, \infty), \leq)$ associated with $X$.

(c) Find the transition density $Q$ of the random walk on $([0, \infty), +)$ associated with $X$.

✓✓

For the last results of this section, suppose again that $X$ has density function $f$ and that $\mathbf{X} = (X_1, X_2, \ldots)$ is the random walk on $(S, \cdot)$ associated with $X$.

**Theorem 5.1.** For $n \in \mathbb{N}_+$,

(a) $(X_1, X_2, \ldots, X_n)$ has density function $g_n$ given by

$$g_n(x_1, x_2, \ldots, x_n) = f(x_1)f(x_1^{-1}x_2)\cdots f(x_{n-1}^{-1}, x_n), \quad x_1 \rightarrow x_2 \rightarrow \cdots \rightarrow x_n$$

(b) $X_n$ has density function $f^{*n}$, the $n$-fold convolution power of $f$.

**Corollary 5.4.** Let $\mathcal{N} = \{N_A : A \in \mathcal{F}\}$ denote the point process associated with $\mathbf{X}$. Then

$$E(N_A) = \sum_{n=1}^{\infty} P(X_n \in A) = \sum_{n=1}^{\infty} \int_A f^{*n}(x) d\lambda(x) = \int_A \left[ \sum_{n=1}^{\infty} f^{*n}(x) \right] d\lambda(x)$$

**Proposition 5.4.** If $X$ has a compound Poisson distribution then $X_n$ has a compound Poisson distribution, with the same compounded distribution, for $n \in \mathbb{N}_+$. 

✓✓
Proof. For \( n \in \mathbb{N}_+ \) we can write \( X_n = \prod_{i=1}^n U_i \) where \( U = (U_1, U_2, \ldots) \) is a sequence of independent variables, each with the distribution of \( X \). Since the distribution of \( X \) is compound Poisson, we can write \( U_i = \prod_{j=1}^{N_i} V_{i,j} \) where \( V_i = (V_{i,1}, V_{i,2}, \ldots) \) is a sequence of independent, identically distributed variables and where \( N_i \) is independent of \( V_i \) has the Poisson distribution with parameter \( \lambda_i \in (0, \infty) \). Since the sequence \( U \) is independent and identically distributed, we can take the collection of random variables \( \{V_{i,j} : i \in \mathbb{N}_+, j \in \mathbb{N}_+\} \) to be independent and identically distributed, and \( (N_1, N_2, \ldots) \) independent. By re-indexing the variables we can write

\[
X = \prod_{i=1}^n \prod_{j=1}^{N_i} V_{i,j} = \prod_{k=1}^N W_k
\]

where \( W = (W_1, W_2, \ldots) \) is a sequence of independent variables, each with the common distribution of \( U_{i,j} \) and where \( N = \sum_{i=1}^n N_i \) has the Poisson distribution with parameter \( \sum_{i=1}^n \lambda_i \).

5.3 Exponential and Memoryless Distributions

In the semigroup setting (and particularly in case of a positive semigroup), there are natural concepts for exponential and memoryless distributions. Most characterizations of the exponential distribution (and its generalizations) in Euclidean spaces are based on the equivalence of the time-shifted distribution with the exponential and memoryless distributions. Most characterizations of the exponential distribution (and its generalizations) in Euclidean spaces are based on the equivalence of the time-shifted distribution with the original distribution, in some sense. So suppose that \( X \) is a random variable supported by \( (S, \cdot) \) with right probability function \( F \). In the setting of a semigroup, exponential properties take the form

\[
P(X \in xA \mid X \in xS) = P(X \in A)
\]

or equivalently

\[
P(X \in xA) = P(X \in xS)P(X \in A) = F(x)P(X \in A)
\]

for certain \( x \in S \) and \( A \in \mathcal{S} \).

**Definition 5.4.** For the semigroup \((S, \cdot)\),

(a) \( X \) has an **exponential distribution** if \( P(X \in xA) = F(x)P(X \in A) \) for \( x \in S \) and \( A \in \mathcal{S} \).

(b) \( X \) has a **memoryless distribution** if \( F(xy) = F(x)F(y) \) for \( x, y \in S \).

Clearly an exponential distribution is memoryless. If we take \( A = yS \) in part (a) we have

\[
F(xy) = P[X \in (xy)S] = P[X \in x(yS)] = F(x)P(X \in yS) = F(x)F(y)
\]

In the case of a positive semigroup, so that the associated relation is a partial order \( \preceq \), the exponential property has the more familiar form

\[
P(x^{-1}X \in A \mid X \succeq x) = P(X \in A); \quad x \in S, A \in \mathcal{S}
\]

and similarly, the memoryless property has the more familiar form

\[
P(X \succeq xy \mid X \succeq x) = P(X \succeq y); \quad x, y \in S
\]

The following definition gives the abstract version of the **new better than used** and **new worse than used** properties.

**Definition 5.5.** For the semigroup \((S, \cdot)\),

(a) \( X \) is **NBU** if \( F(xy) \leq F(x)F(y) \) for \( x, y \in S \).

(b) \( X \) is **NWU** if \( F(xy) \geq F(x)F(y) \) for \( x, y \in S \).

So more explicitly, the NBU and NWU properties are, respectively

\[
P(xy \to X \mid x \to X) \leq P(y \to X), \quad x, y \in S
\]

\[
P(xy \to X \mid x \to X) \geq P(y \to X), \quad x, y \in S
\]
Once again, in the case of a positive semigroup, with a partial order $\preceq$ as the relation, these properties take the more recognizable form

\[
\begin{align*}
\mathbb{P}(X \geq xy \mid X \geq x) &\leq \mathbb{P}(X \geq y), \quad x, y \in S \\
\mathbb{P}(X \geq xy \mid X \geq x) &\geq \mathbb{P}(X \geq y), \quad x, y \in S
\end{align*}
\]

We will concentrate on the memoryless and exponential properties in this text.

**Example 5.1.** Suppose that $(S, \cdot)$ is a right trivial semigroup. That is, $xy = y$ for $x, y \in S$, and hence it follows that $x A = A$ for $x \in S$ and $A \in \mathcal{S}$. If $X$ is a random variable with values in $S$ then

\[
\mathbb{P}(X \in x A) = \mathbb{P}(X \in A) = \mathbb{P}(X \in S) \mathbb{P}(X \in A) = \mathbb{P}(X \in x S) \mathbb{P}(X \in A), \quad x \in S, A \in \mathcal{S}
\]

Thus every probability distribution is exponential for a right trivial semigroup, just as every $\sigma$-finite measure is left invariant.

**Theorem 5.2.** Suppose again that $X$ is a random variable supported by $(S, \cdot)$ and with right probability function $F$. Define

\[S_X = \{x \in S : \mathbb{P}(X \in x A) = F(x) \mathbb{P}(X \in A) \text{ for all } A \in \mathcal{S}\}\]

If $S_X \neq \emptyset$ then $S_X$ is a complete sub-semigroup of $(S, \cdot)$.

**Proof.** We first show closure. Suppose that $x, y \in S_X$. Then for $A \in \mathcal{S}$,

\[
\mathbb{P}[X \in (xy) A] = \mathbb{P}[X \in x(yA)] = F(x) \mathbb{P}(X \in y A) = F(x) F(y) \mathbb{P}(X \in A)
\]

In particular, letting $A = S$ we have $F(xy) = F(x) F(y)$, so substituting back we have

\[
\mathbb{P}[X \in (xy) A] = F(xy) \mathbb{P}(X \in A)
\]

and so $xy \in S_X$. Next we show completeness. Suppose that $x, y \in S_X$ and that $x \rightarrow y$ so that $y = xu$ for some $u \in S$. We need to show that $u = x^{-1} y \in S_X$. Let $A \in \mathcal{S}$. First, since $x \in S_X$ we have

\[
\mathbb{P}[X \in (xu) A] = \mathbb{P}[X \in x(uA)] = F(x) \mathbb{P}(X \in uA)
\]

On the other hand, since $y = xu \in S_X$ we have

\[
\mathbb{P}[X \in (xu) A] = F(xu) \mathbb{P}(X \in A)
\]

Again since $x \in S_X$ we have

\[
F(xu) = \mathbb{P}[X \in (xu) S] = \mathbb{P}[X \in x(uS)] = F(x) \mathbb{P}(X \in uS) = F(x) F(u)
\]

Combining the displayed equations we have

\[
F(x) \mathbb{P}(X \in u A) = F(x) F(u) \mathbb{P}(X \in A)
\]

Since $F(x) > 0$ we have $\mathbb{P}(X \in u A) = F(u) \mathbb{P}(X \in A)$, so $u \in S_X$. \hfill \square

In the case that $(S, \cdot)$ is a positive semigroup, note that the identity element $e \in S_X$ so $(S_X, \cdot)$ is a complete positive sub-semigroup. Of course in general, $S_X$ may be empty, or in the case of a positive semigroup we could have $S_X = \{e\}$. These cases aside, every distribution satisfies the exponential property on some complete sub-semigroup.

**Theorem 5.3.** Suppose again that $X$ is a random variable with values in $S$ with right probability function $F$. Define

\[\mathcal{S}_X = \{A \in \mathcal{S} : \mathbb{P}(X \in x A) = F(x) \mathbb{P}(X \in A) \text{ for all } x \in S\}\]

Then $\mathcal{S}_X$ is closed under countable disjoint unions, proper differences, countable increasing unions, and countable decreasing intersections (and hence is a monotone class).
Proof. Let \((A_1, A_2, \ldots)\) be a sequence of disjoint sets in \(\mathcal{F}_X\) and let \(x \in S\). Since the semigroup is left-cancellative, \((xA_1, xA_2, \ldots)\) is a disjoint sequence and
\[
\mathbb{P} \left( X \in x \bigcup_{i=1}^{\infty} A_i \right) = \mathbb{P} \left( X \in \bigcup_{i=1}^{\infty} xA_i \right) = \sum_{i=1}^{\infty} \mathbb{P}(X \in xA_i) = \sum_{i=1}^{\infty} F(x)\mathbb{P}(X \in A_i)
\]
\[
= F(x) \sum_{i=1}^{\infty} \mathbb{P}(X \in A_i) = F(x)\mathbb{P} \left( X \in \bigcup_{i=1}^{\infty} A_i \right)
\]
Hence \(\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_X\). Next let \(A, B \in \mathcal{F}_X\) with \(A \subseteq B\) and let \(x \in S\). Then \(xA \subseteq xB\) and since the semigroup is left-cancellative, \(x(B-A) = xB - xA\). Hence
\[
\mathbb{P}(X \in x(B - A)) = \mathbb{P}(X \in (xB - xA)) = \mathbb{P}(X \in xB) - \mathbb{P}(X \in xA)
\]
\[
= F(x)\mathbb{P}(X \in A) - F(x)\mathbb{P}(X \in B) = F(x)[\mathbb{P}(X \in B) - \mathbb{P}(X \in A)]
\]
\[
= F(x)\mathbb{P}(X \in B - A)
\]
Hence \(B - A \in \mathcal{F}_X\). Of course \(S \in \mathcal{F}_X\) and hence the other results follow.

Our next result and its corollary give expected value characterizations of the exponential property. If \(X\) is a random variable with values in \(S\) and \(g : S \to \mathbb{R}\) is measurable, recall that the expected value \(\mathbb{E}[g(X)]\) is well defined if either \(g\) is nonnegative (although \(\infty\) is a possible value) or \(\mathbb{E}[|g(X)|] < \infty\).

**Theorem 5.4.** Suppose again that \(X\) is a random variables supported by \((S, \cdot)\) and right probability function \(F\). If \(X\) has an exponential distribution for \((S, \cdot)\) then
\[
\mathbb{E}[g(x^{-1}X), X \in xS] = F(x) \mathbb{E}[g(X)]
\]
for \(x \in S\) and measurable \(g : S \to \mathbb{R}\), assuming that the expected values are defined.

**Proof.** Recall that if \(x \in S\) and \(y \in xS\) then \(x^{-1}y\) is the unique element in \(S\) satisfying \(x(x^{-1}y) = y\). Moreover, \(x^{-1}y \in A\) if and only if \(y \in xA\) for \(A \in \mathcal{F}\). Suppose first that \(g = 1_A\) for \(A \in \mathcal{F}\). Since \(X\) has an exponential distribution for \((S, \cdot)\),
\[
\mathbb{E}[g(x^{-1}X), X \in xS] = \mathbb{P}(x^{-1}X \in A, X \in xS) = \mathbb{P}(X \in xA) = F(x)\mathbb{P}(X \in A)
\]
On the other hand \(F(x) \mathbb{E}[g(X)] = F(x)\mathbb{P}(X \in A)\). Hence (5.1) holds for indicator random variables. A standard bootstrapping argument then implies that the (5.1) holds for simple functions, for nonnegative functions, and finally whenever the expected values make sense.

Conversely (as in the proof), if (5.1) holds for functions of the form \(g = 1_A\) where \(A \in \mathcal{F}\), then \(X\) has an exponential distribution.

**Corollary 5.5.** Suppose again that \(X\) is a random variable supported by \((S, \cdot)\) and with right probability function \(F\). If \(X\) has an exponential distribution then
\[
\mathbb{E}[h(X), X \in xA] = F(x) \mathbb{E}[h(xX), X \in A]
\]
for \(x \in S\), \(A \in \mathcal{F}\), and measurable \(h : S \to \mathbb{R}\), assuming that expected values are defined.

**Proof.** Let \(x \in S\), \(A \in \mathcal{F}\), and \(h : S \to \mathbb{R}\) measurable. Define \(g : S \to \mathbb{R}\) by \(g(y) = 1_A(y)h(xy)\) for \(y \in S\). Then \(g\) is also measurable so Theorem 5.4 applies. But for \(y \in xS\),
\[
g(x^{-1}y) = 1_A(x^{-1}y)h(y) = 1_{xA}(y)h(y)
\]
since \(x^{-1}y \in A\) if and only if \(y \in xA\). Hence
\[
\mathbb{E}[g(x^{-1}X), X \in xS] = \mathbb{E}[1_{xA}(X)h(X), X \in xS] = \mathbb{E}[h(X), X \in xA]
\]
On the other hand,
\[
F(x) \mathbb{E}[g(X)] = F(x) \mathbb{E}[1_A(X)h(xX)] = F(x) \mathbb{E}[h(xX), X \in A]
\]
So far, we have not needed to refer to a reference measure \( \lambda \) on \( (S, \mathcal{S}) \) or a density function of \( X \) with respect to \( \lambda \), as we did for constant rate distributions. The following theorem bridges the gap and gives one of the main characterization of exponential distributions.

**Theorem 5.5.** Suppose again that \( X \) is a random variable supported by \( (S, \cdot) \). Then \( X \) has an exponential distribution for \( (S, \cdot) \) if and only if \( X \) is memoryless for \( (S, \cdot) \) and has constant rate for \( (S, \rightarrow) \) with respect to a left-invariant measure on \( (S, \mathcal{S}) \).

**Proof.** Let \( f \) denote the right probability function of \( X \). Suppose first that \( X \) has an exponential distribution. Then as noted above, the distribution is also memoryless. Now let \( \mu \) be the \( \sigma \)-finite positive measure defined by

\[
\mu(A) = \mathbb{E} \left( \frac{1}{F(X)}, X \in A \right), \quad A \in \mathcal{S}
\]

Then by Theorem 4.9, \( X \) has constant rate 1 with respect to \( \mu \):

\[
\mathbb{P}(X \in A) = \int_A f(x) d\mu(x), \quad A \in \mathcal{S}
\]

Let \( x \in S \) and \( A \in \mathcal{S} \). By the integral version of the exponential property in Corollary 5.5 and by the memoryless property,

\[
\mu(xA) = \mathbb{E} \left( \frac{1}{F(X)}, X \in xA \right) = f(x) \mathbb{E} \left( \frac{1}{F(xX)}, X \in A \right) = f(x) \mathbb{E} \left( \frac{1}{F(X)}, X \in A \right) = \mu(A)
\]

so \( \mu \) is left invariant. Conversely, suppose that the distribution of \( X \) is memoryless and has constant rate \( \alpha \in (0, \infty) \) with respect to a left-invariant measure \( \lambda \). Thus \( f = \alpha F \) is a density function of \( X \) with respect to \( \lambda \). Let \( x \in S \) and \( A \in \mathcal{S} \). Using the memoryless property and the integral version of left invariance in Corollary 3.6,

\[
\mathbb{P}(X \in xA) = \int_{xA} \alpha F(y) d\lambda(y) = \int_A \alpha F(xz) d\lambda(z) = \int_A \alpha F(x) F(z) d\lambda(z) = F(x) \int_A \alpha F(z) d\lambda(z) = F(x) \mathbb{P}(X \in A)
\]

Hence \( X \) has an exponential distribution. \( \square \)

In particular, if \( (S, \cdot) \) has an exponential distribution then \( (S, \cdot) \) has a left-invariant measure, not surprising since the existence of an exponential distribution requires somewhat more of \( (S, \cdot) \) than the existence of a left-invariant measure. On the other hand, if \( (S, \cdot) \) has a left-invariant measure \( \lambda \) that is unique up to multiplication by positive constants, then \( X \) has an exponential distribution for \( (S, \cdot) \) if and only if \( X \) is memoryless and constant rate with respect to \( \lambda \). In particular, this statement applies to discrete positive semigroups, with counting measure \# as the left invariant measure.

**Corollary 5.6.** Suppose that \( (S, \cdot) \) is a discrete positive semigroup with associated partial order \( \preceq \). Then \( X \) has an exponential distribution for \( (S, \cdot) \) if and only if \( X \) is memoryless for \( (S, \cdot) \) and has constant rate \( \alpha \) for \( (S, \preceq) \), for some \( \alpha \in (0, 1) \).

**Proof.** Note that \( \mathbb{P}(X = e) = \alpha \mathbb{P}(X \geq e) = \alpha \) so we must have \( \alpha \in (0, 1) \). \( \square \)

**Theorem 5.6.** Suppose that \( F : S \to (0, 1] \) is measurable. Then \( F \) is the right probability function of an exponential distribution for \( (S, \cdot) \) that has constant rate with respect to a left-invariant measure \( \lambda \) if and only if

\[
F(xy) = F(x) F(y), \quad x, y \in S
\]

\[
\int_S F(x) d\lambda(x) < \infty
\]
Proof. Suppose first that \( F \) is the right probability function of an exponential distribution for \((S, \cdot)\) that has constant rate \( \alpha \in (0, \infty) \). By Theorem 5.5 the distribution is memoryless, so (5.2) holds. Also \( \alpha F \) is a probability density function so
\[
\int_S F(x) d\lambda(x) = \frac{1}{\alpha} < \infty
\]
and hence (5.3) holds.

Conversely, suppose that (5.2) and (5.3) hold. Let \( f = \alpha F \) where \( \alpha = 1/\int_S F(x) d\lambda(x) \). Then by (5.3), \( f \) is a probability density function. Let \( X \) be a random variable with density \( f \), and let \( x \in S \) and \( A \in \mathcal{A} \). Using (5.2) and the integral version of the left invariance property in Corollary 3.6,
\[
\mathbb{P}(X \in xA) = \int_{xA} f(y) d\lambda(y) = \int_{xA} \alpha F(y) d\lambda(y) = \int_A \alpha F(xz) d\lambda(z) = F(x) \int_A f(z) d\lambda(z) = F(x) \mathbb{P}(X \in A)
\]
Letting \( A = S \) we see that \( F \) is the right probability function of \( X \), and so it then follows that \( X \) has an exponential distribution with rate \( \alpha \).

If \( \lambda \) is the unique left invariant measure for \((S, \cdot)\), up to multiplication by positive constants, then this theorem gives a method for finding all exponential distributions. It also follows that the memoryless property almost implies the constant rate property (and hence the full exponential property). More specifically, if \( F \) is a right probability function satisfying the conditions in Theorem 5.6, then \( f = \alpha F \) is the probability density function of an exponential distribution with right probability function \( F \) (where again, \( \alpha = 1/\int_S F(x) d\lambda(x) \)). But in general, there may be other probability density functions with same right probability function \( F \) that are not multiples of \( F \) (and hence do not have constant rate). It may also be possible for \( F \) to satisfy (5.2) but with \( \int_S F(x) d\lambda(x) = \infty \). But to emphasize, we do have the following:

**Theorem 5.7.** If \((S, \cdot)\) is a semigroup where right probability functions uniquely determine distributions, then a distribution is exponential if and only if it is memoryless.

Section 17.7 gives an example of a discrete, positive semigroup where the right probability function does not determine the distribution and where there are memoryless distributions that are not exponential. The following trivial example may also help.

**Example 5.2.** Suppose that \((S, \cdot)\) is a right trivial semigroup, so that \( xy = y \) for \( x, y \in S \). The corresponding relation is the complete relation \( \equiv \) so that \( x \equiv y \) for every \( (x, y) \in S^2 \). Recall that every measure is invariant for \((S, \cdot)\), every probability distribution on \( S \) is exponential for \((S, \cdot)\), and that the probability function for \((S, \equiv)\) of every distribution is simply the constant function 1. Suppose that \( \lambda \) is the reference measure on \((S, \mathcal{A})\). If \( \lambda(S) < \infty \) then the only distribution with constant rate for \((S, \equiv)\) is the uniform distribution (with rate \( 1/\lambda(S) \)). If \( \lambda(S) = \infty \), there are no constant rate distributions. So this is a trivial example of a semigroup that has exponential (and hence memoryless) distributions that do not have constant rate. We can also view this example through the lens of Theorem 5.6. If \( F(xy) = F(x)F(y) \) then \( F(y) = F(x)F(y) \) for \( x, y \in S \). From our support assumption, \( F(y) > 0 \) so \( F(x) = 1 \) for \( x \in S \). Also, \( \int_S F(x) d\lambda(x) = \lambda(S) \). So condition (5.3) holds if and only if \( \lambda \) is a finite measure, in which case the corresponding constant rate distribution is simply the uniform distribution on \( S \) (with respect to \( \lambda \))
\[
\mathbb{P}(X \in A) = \frac{\lambda(A)}{\lambda(S)}, \quad A \in \mathcal{A}
\]
On the other hand, if \( \lambda(S) = \infty \) then there is no distribution that has constant rate with respect to \( \lambda \). So to summarize, a distribution on \( S \) (which is necessarily exponential for \((S, \cdot)\)) has constant rate with respect to a measure \( \lambda \) (which is necessarily left invariant) if and only if \( \lambda \) is a finite measure, in which case the distribution is uniform. This statement is not as restrictive as it might seem. The canonical measure associated with \( X \) is simply the distribution of \( X \):
\[
\mathbb{E} \left[ \frac{1}{F(X)} ; X \in A \right] = \mathbb{P}(X \in A), \quad A \in \mathcal{A}
\]
and trivially, \( X \) has the uniform distribution with respect to itself:
\[
\mathbb{P}(X \in A) = \frac{\mathbb{P}(X \in A)}{\mathbb{P}(X \in S)}, \quad A \in \mathcal{A}
\]
Conversely, the constant rate property does not imply the memoryless property. The following general example shows that mixtures of distinct exponential distributions with the same constant rate will still have the constant rate property, but not the memoryless property. The free semigroup studied in Chapter gives a specific example where there are different exponential distributions with the same rate.

**Example 5.3.** Suppose that \((S, \cdot)\) is a semigroup with a fixed left-invariant measure \(\lambda\). Suppose that \(F\) and \(G\) are right probability functions for distinct exponential distributions for \((S, \cdot)\), each having constant rate \(\alpha \in (0, \infty)\) with respect to \(\lambda\). Let \(p \in (0,1)\) and \(H = pF + (1-p)G\). Then \(H\) is also the right probability function for a distribution with constant rate \(\alpha\). The distributions corresponding to \(F\) and \(G\) are memoryless, but not the distribution corresponding to \(H\):

\[
H(xy) = pF(xy) + (1-p)G(xy) = pF(x)F(y) + (1-p)G(x)G(y), \quad x, y \in S
\]

while

\[
H(x)H(y) = [pF(x) + (1-p)G(x)][pF(y) + (1-p)G(y)] = p^2F(x)F(y) + p(1-p)[F(x)G(y) + F(y)G(x)] + (1-p)^2F(y)G(y), \quad x, y \in S
\]

**Corollary 5.7.** Suppose that \(F\) is the right probability function of an exponential distribution for \((S, \cdot)\) that has constant rate with respect to a left-invariant measure \(\lambda\). If \(m \in (0, \infty)\) and

\[
1/\alpha_m := \int_S F^m(x)d\lambda(x) < \infty
\]

(in particular if \(m \geq 1\), then \(F^m\) is the right probability function of an exponential distribution for \((S, \cdot)\) that has rate \(\alpha_m\) with respect to \(\lambda\).

**Proof.** Clearly \(F^m(xy) = F^m(x)F^m(y)\) for \(x, y \in S\). Thus the result follows immediately from Theorem 5.6.

**Theorem 5.8.** Suppose that \((S, \cdot)\) is a positive semigroup and that \(f\) is a probability density function with respect to a left-invariant measure \(\lambda\) satisfying

\[
f(x)f(y) = G(xy), \quad x, y \in S
\]

for some measurable function \(G : S \to (0, \infty)\). Then \(f\) is the density of an exponential distribution for \((S, \cdot)\).

**Proof.** Let \(e\) denote the identity element of \((S, \cdot)\). Letting \(y = e\) in the displayed equation gives \(G(x) = \alpha f(x)\) where \(\alpha = f(e) \in (0, \infty)\). Let \(F\) denote the right probability function of \(f\) for \((S, \cdot)\). Then using the integral version of the left-invariance property in Corollary 3.6,

\[
F(x) = \int_{xS} f(y)d\lambda(y) = \int_{xS} 1/\alpha G(y)d\lambda(y) = \int_S G(xu)d\lambda(u)
\]

\[
= \frac{1}{\alpha} \int_S f(x)f(u)d\lambda(u) = \frac{1}{\alpha} f(x), \quad x \in S
\]

Thus the distribution has constant rate \(\alpha\). Finally,

\[
F(xy) = \frac{1}{\alpha} f(xy) = \frac{1}{\alpha^2} G(xy) = \frac{1}{\alpha^2} f(x)f(y) = F(x)F(y)
\]

so the distribution is memoryless. Hence \(f\) is the density of an exponential distribution by Theorem 5.5.

Suppose that \(X\) is a random variable supported by \((S, \cdot)\). As noted earlier, there are two random walks associated with \(X\)—one on the graph \((S, \to)\) and one on the semigroup \((S, \cdot)\). We can now answer the question of when these are the same:

**Theorem 5.9.** The random walk on \((S, \to)\) associated with \(X\) is the same as the random walk on \((S, \cdot)\) associated with \(X\) if and only if \(X\) has an exponential distribution for \((S, \cdot)\).
Proof. By definition, the two random walks are the same if and only if the conditional distribution of $X$ given $x \to X$ is the same as the distribution of $xX$ for $x \in S$. Both distributions are concentrated on $xS$. Let $x \in S$ and $A \in \mathcal{S}$ so that $xA \subseteq xS$. Then

$$P(X \in xA \mid x \to X) = P(X \in xA \mid X \in xS) = \frac{P(X \in xA)}{P(X \in xS)}$$

and $P(xX \in xA) = P(X \in A)$. Hence if the random walks are the same then

$$\frac{P(X \in xA)}{P(X \in xS)} = P(X \in A), \quad x \in S, A \in \mathcal{S}$$

and hence $X$ has an exponential distribution. Conversely, suppose that $X$ has an exponential distribution. Let $x \in S$ and $B \in \mathcal{S}$ with $B \subseteq xS$. Then $B = xA$ where $A = x^{-1}B$. By the exponential property,

$$P(X \in xA \mid X \in xS) = P(X \in A)$$

The left side is $P(X \in B \mid x \to X)$ while the right side is $P(xX \in B)$. Hence the random walks are the same.

For another proof, suppose that $X$ has probability density function $f$ with respect to a fixed left-invariant reference measure $\lambda$ for $(S, \cdot)$ and has right probability function of $F$. Recall that the random walk on $(S, \to)$ associated with $X$ has transition density $P$ given by $P(x, y) = f(y)/F(x)$ for $x \in S$ and $y \in xS$. The random walk on $(S, \cdot)$ associated with $X$ has transition density $Q$ given by $Q(x, y) = f(x^{-1}y)$ for $x \in S$ and $y \in xS$. Suppose that $X$ has an exponential distribution with constant rate $\alpha \in (0, \infty)$ with respect to $\lambda$. Let $x \in S$ and $y \in xS$ (so that $x \to y$). By definition, $x(x^{-1}y) = y$. Hence by the memoryless property

$$F(y) = F[x(x^{-1}y)] = F(x)F(x^{-1}y)$$

so $F(x^{-1}y) = F(y)/F(x)$. Hence

$$Q(x, y) = f(x^{-1}y) = \frac{1}{\alpha}F(x^{-1}y) = \frac{1}{\alpha}F(y) = \frac{f(y)}{F(x)} = P(x, y)$$

Conversely, suppose that $P = Q$ so that

$$\frac{f(y)}{F(x)} = f(x^{-1}y), \quad x \in S, y \in xS$$

Then $f(y) = f(x^{-1}y)F(x)$ for $x \in S$ and $y \in xS$, or equivalently, $f(xt) = f(t)F(x)$ for $x, t \in S$. If $A \in \mathcal{S}$ then by the left invariance property,

$$P(X \in xA) = \int_{xA} f(y) d\lambda(y) = \int_A f(xt) d\lambda(t) = \int_A f(t)F(x) d\lambda(t) = F(x)P(X \in A)$$

Hence $X$ has an exponential distribution.

\section{Conditional Distributions}

A number of characterizations of the standard exponential distribution deal with conditional distributions in various ways. In this section we assume that $(S, \cdot)$ is a semigroup.

\textbf{Theorem 5.10}. Suppose that $X$ and $Y$ are independent random variables supported by $(S, \cdot)$, and with right probability functions $F$ and $G$. Suppose also that $X$ has an exponential distribution and that $Y$ has a memoryless distribution. Then the conditional distribution of $X$ given $Y \in xS$ is exponential, with right probability function $FG$.

\textbf{Proof}. First, since both distributions are memoryless, we have

$$(FG)(xy) = F(xy)G(xy) = F(x)F(y)G(x)G(y)$$

$$= [F(x)G(x)][F(y)G(y)] = (FG)(x)(FG)(y), \quad x, y \in S$$
Next, $X$ has constant rate $\alpha \in (0, \infty)$ with respect to a left-invariant measure $\lambda$. Since $G : S \to (0, 1]$ we have
\[
\frac{1}{\gamma} := \int_S F(x)G(x)d\lambda(x) \leq \int_S F(x)d\lambda(x) = \frac{1}{\alpha} < \infty
\]
From Theorem 5.6, it follows that $FG$ is the right probability function of an exponential distribution that has constant rate $\gamma$ with respect to $\lambda$. It remains to show that this distribution is the conditional distribution of $X$ given $Y \in XS$. Towards this end, note that
\[
\mathbb{P}(Y \in XS) = \mathbb{E}[\mathbb{P}(Y \in XS \mid X)] = \mathbb{E}[G(X)]
\]
\[
= \int_X G(x)\alpha F(x)d\lambda(x) = \alpha \int_S G(x)F(x)d\lambda(x) = \frac{\alpha}{\gamma}
\]

Next, if $A \in \mathcal{S}$, then
\[
\mathbb{P}(X \in A, Y \in XS) = \mathbb{E}[\mathbb{P}(X \in A, Y \in XS \mid X)]
\]
\[
= \mathbb{E}[G(X), X \in A] = \int_A G(x)\alpha F(x)d\lambda(x)
\]
and therefore
\[
\mathbb{P}(X \in A \mid Y \in XS) = \int_A \frac{\gamma F(x)G(x)d\lambda(x)}{\mathbb{P}(Y \in XS)}
\]
So the conditional density of $X$ given $Y \in XS$ is $\gamma FG$. Using the integral version of left invariance in Corollary 3.6,
\[
\mathbb{P}(X \in xS \mid Y \in XS) = \gamma \int_x S F(y)G(y)d\lambda(y) = \gamma \int_S F(xz)G(z)d\lambda(z)
\]
\[
= F(x)G(x) \int_x S \gamma F(z)G(z)d\lambda(z) = F(x)G(x), \quad x \in S
\]
so $FG$ is the right probability function of $X$ given $Y \in XS$.

In the context of the previous theorem, suppose that $X$ and $Y$ are independent variables supported by $(S, \cdot)$, and that each has an exponential distribution. Then the conditional distribution of $X$ given $Y \in XS$ and the conditional distribution of $Y$ given $X \in XS$ are the same—exponential with right probability function $FG$.

**Theorem 5.11.** Suppose that $X$ and $Y$ are independent random variables supported by $(S, \cdot)$, and that the distribution of $Y$ is exponential with right probability function $G$.

(a) The variables $X$ and $X^{-1}Y$ are conditionally independent given $Y \in XS$.

(b) The conditional distribution of $X^{-1}Y$ given $Y \in XS$ is the same as the distribution of $Y$.

(c) The conditional distribution of $X$ given $Y \in XS$ is defined by
\[
\mathbb{P}(X \in A \mid Y \in XS) = \frac{\mathbb{E}[G(X), X \in A]}{\mathbb{E}[G(X)]}, \quad A \in \mathcal{S}
\]

**Proof.** As in the previous theorem, let
\[
\frac{1}{\gamma} = \mathbb{P}(Y \in XS) = \mathbb{E}[\mathbb{P}(Y \in XS \mid X)] = \mathbb{E}[G(X)]
\]
Let $A, B \in \mathcal{S}$. Using the exponential property of $Y$,
\[
\mathbb{P}(X \in A, X^{-1}Y \in B \mid Y \in XS) = \frac{\mathbb{P}(X \in A, X^{-1}Y \in B, Y \in XS)}{\mathbb{P}(Y \in XS)}
\]
\[
= \gamma \mathbb{P}(X \in A, Y \in XB) = \gamma \mathbb{E}[\mathbb{P}(X \in A, Y \in XB \mid X)]
\]
\[
= \gamma \mathbb{E}[\mathbb{P}(Y \in XB \mid X), X \in A]
\]
\[
= \gamma \mathbb{E}[\mathbb{P}(Y \in XS \mid X)\mathbb{P}(Y \in B \mid X), X \in A]
\]
\[
= \gamma \mathbb{E}[G(X)\mathbb{P}(Y \in B), X \in A] = \gamma \mathbb{P}(Y \in B) \mathbb{E}[G(X), X \in A]
\]
CHAPTER 5. PROBABILITY ON SEMIGROUPS

52

It follows that $X$ and $X^{-1}Y$ are conditionally independent given $Y \in XS$. Letting $A = S$,

$$\mathbb{P}(X^{-1}Y \in B \mid Y \in XS) = \mathbb{P}(Y \in B), \quad B \in \mathcal{F}$$

Letting $B = S$,

$$\mathbb{P}(X \in A \mid Y \in XS) = \frac{\mathbb{E}[G(X), X \in A]}{\mathbb{E}[G(X)]}, \quad A \in \mathcal{F}$$

\[ \square \]

Suppose now that $(T, \cdot)$ is a measurable sub-semigroup of $(S, \cdot)$. Recall that the underlying measurable space for $(T, \cdot)$ is $(T, \mathcal{F})$ where $\mathcal{F} = \{ A \in \mathcal{F} : A \subseteq T \}$.

**Theorem 5.12.** Suppose that $X$ has an exponential distribution for $(S, \cdot)$ with right probability function $F$ and that $\mathbb{P}(X \in T) > 0$. Then

(a) The right probability function of $X$ given $X \in T$ for $(T, \cdot)$ is the restriction of $F$ to $T$.

(b) The distribution of $X$ given $X \in T$ is exponential for $(T, \cdot)$.

(c) If $X$ has constant rate $\alpha \in (0, \infty)$ for $(S, \cdot)$ with respect to a left-invariant measure $\lambda$, then the distribution of $X$ given $X \in T$ has constant rate $\alpha/\mathbb{P}(X \in T)$ for $(T, \cdot)$ with respect to the restriction of $\lambda$ to $\mathcal{F}$.

**Proof.** The proofs are simple.

(a) Let $F_T$ denote the right probability function of $X$ given $X \in T$ relative to $(T, \cdot)$. That is,

$$F_T(x) = \mathbb{P}(X \in xT \mid X \in T), \quad x \in T$$

Since $xT \subseteq T$ and by the exponential property of $X$ we have

$$F_T(x) = \frac{\mathbb{P}(X \in xT)}{\mathbb{P}(X \in T)} = \frac{\mathbb{P}(X \in xS)\mathbb{P}(X \in T)}{\mathbb{P}(X \in T)} = \mathbb{P}(X \in xS) = F(x), \quad x \in T$$

(b) Let $A \in \mathcal{F}$ and $x \in T$. Again since $xA \subseteq T$ and by the exponential property of $X$ we have

$$\mathbb{P}(X \in xA \mid X \in T) = \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(X \in T)} = \frac{\mathbb{P}(X \in xS)\mathbb{P}(X \in A)}{\mathbb{P}(X \in T)} = F_T(x)\mathbb{P}(X \in A \mid X \in T)$$

(c) Let $f = \alpha F$ on $S$ so that $f$ is a density function of $X$. The density function $f_T$ of $X$ given $X \in T$ is

$$f_T(x) = f(x)/\mathbb{P}(X \in T)$$

for $x \in T$. Hence

$$f_T(x) = \frac{f(x)}{\mathbb{P}(X \in T)} = \frac{\alpha F(x)}{\mathbb{P}(X \in T)} = \frac{\alpha}{\mathbb{P}(X \in T)} F(x), \quad x \in T$$

\[ \square \]

We will extend this result in Chapter 9 on quotient spaces. For the corollary that follows, suppose that $(S, \cdot)$ is a discrete positive semigroup with identity element $e$, and let $S_+ = \{ x \in S : x \neq e \}$. Then the sub-semigroup $(S_+, \cdot)$ is a strict positive semigroup. The relation $\preceq$ associated with $(S, \cdot)$ is a partial order, and the relation $\preceq$ associated with $(S_+, \cdot)$ is the corresponding strict partial order. Of course, counting measure $\#$ is left invariant for $(S, \cdot)$, and is the unique such measure, up to multiplication by positive constants. Suppose now that $X$ is a random variable supported by $(S, \cdot)$.

**Corollary 5.8.** If $X$ has an exponential distribution for $(S, \cdot)$ with constant rate $\alpha \in (0, \infty)$ then the distribution of $X$ given $X \in S_+$ is exponential for $(S_+, \cdot)$ with constant rate $\alpha/(1 - \alpha)$.

**Proof.** That $X$ has an exponential distribution follows from Theorem 5.12. Moreover, $\mathbb{P}(X \in S_+) = 1 - f(e) = 1 - \alpha$.

\[ \square \]

We return to the general setting of a semigroup $(S, \cdot)$ with a measurable sub-semigroup $(T, \cdot)$. Suppose that $\lambda$ is a fixed left-invariant measure for $(S, \cdot)$. The following result revisits Theorem 5.12, but with only the constant rate property.
Proposition 5.5. Suppose that $X$ has constant rate $\alpha \in (0, \infty)$ for $(S, \cdot)$ and that $\mathbb{P}(X \in T) > 0$. Then the conditional distribution of $X$ given $X \in T$ has right rate function $r_T$ for $(T, \cdot)$ defined by

$$r_T(x) = \alpha \frac{\mathbb{P}(X \in xS)}{\mathbb{P}(X \in xT)}, \quad x \in T$$

Proof. Let $F$ denote the right probability function of $X$ for $(S, \cdot)$ so that $F(x) = \mathbb{P}(X \in xS)$ for $x \in S$. By assumption, $f = \alpha F$ is a density of $X$. As before, the right probability function $F_T$ of $X$ given $X \in T$ for $(T, \cdot)$ is defined by

$$F_T(x) = \mathbb{P}(X \in xT \mid X \in T) = \frac{\mathbb{P}(X \in xT)}{\mathbb{P}(X \in T)}, \quad x \in T$$

The density function $f_T$ of $X$ given $X \in T$ is defined by

$$f_T(x) = \frac{f(x)}{\mathbb{P}(X \in T)} = \alpha \frac{F(x)}{\mathbb{P}(X \in T)}$$

Hence the right rate function $r_T$ of $X$ given $X \in T$ for $(T, \cdot)$ is defined by

$$r_T(x) = \frac{f_T(x)}{F_T(x)} = \frac{\mathbb{P}(X \in xS)}{\mathbb{P}(X \in xT)}, \quad x \in T$$
Part II

Special Constructions
Chapter 6

Reflexive Completion

In this brief chapter, we study a simple construction that is helpful for the analysis of discrete graphs. To begin, suppose that $(S, \rightarrow)$ is a discrete, irreflexive graph, so in the combinatorial sense, a graph with no loops.

**Definition 6.1.** The reflexive completion of $(S, \rightarrow)$ is the graph $(S, \leftrightarrow)$ where $x \leftrightarrow y$ if and only if $x = y$ or $x \rightarrow y$ for $x, y \in S$.

That is, we simply add a loop at each vertex. Equivalently, we could start with a discrete, reflexive graph $(S, \leftrightarrow)$ and remove the loops to form the irreflexive graph $(S, \rightarrow)$. We might call this reflexive deletion.

Note in particular that if $(S, \prec)$ is a strict partial order graph (anti-reflexive, asymmetric, and transitive) then the reflexive completion is $(S, \preceq)$, the corresponding partial order graph (reflexive, anti-symmetric, and transitive). We will use our usual notation for mathematical objects related to the graphs $(S, \rightarrow)$ and $(S, \leftrightarrow)$, but with the addition of a bar overscript for the latter. The following result connects the left walk functions for $(S, \rightarrow)$ and $(S, \leftrightarrow)$.

**Theorem 6.1.** Suppose that $\gamma_n$ and $\bar{\gamma}_n$ denote the left walk functions of order $n \in \mathbb{N}$, for $(S, \rightarrow)$ and $(S, \leftrightarrow)$ respectively. Then for $x \in S$,

$$\bar{\gamma}_n(x) = \sum_{k=0}^{n} \binom{n}{k} \gamma_k(x)$$

$$\gamma(x) = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} \bar{\gamma}_k(x)$$

**Proof.** Let $n \in \mathbb{N}_+$ and $x \in S$. Consider a walk $x = (x_1, x_2, \ldots, x_k, x_{k+1})$ of length $k$ in $(S, \rightarrow)$, where $k \in \mathbb{N}$, $k \leq n$, and $x_{k+1} = x$ (so the walk ends at $x$). Since $\rightarrow$ is anti-reflexive, $x_{j+1} \neq x_j$ for $j \in \{1, 2, \ldots, k\}$. To construct a walk $y$ of length $n$ in $(S, \leftrightarrow)$ that ends in $x$, we can add $n-k$ loops to the walk $x$. Conversely, every walk of length $n$ in $(S, \leftrightarrow)$ that ends in $x$ must have this form for some $k \leq n$. In the chain $y$, let $a_j$ be the number of repetitions (loops) of $x_j$ for $j \in \{1, 2, \ldots, k+1\}$. Then $a_j$ is a nonnegative integer and $\sum_{j=1}^{k+1} a_j = n - k$. The number of solutions is

$$\binom{(k+1) + (n-k) - 1}{n-k} = \binom{n}{k}$$

By definition, there are $\delta_k(x)$ walks of length $k$ in $(S, \rightarrow)$ that end in $x$, so it follows that the number of walks of length $n$ for $(S, \leftrightarrow)$ that end in $x$ is

$$\bar{\gamma}_n(x) = \sum_{k=0}^{n} \binom{n}{k} \gamma_k(x)$$

The second result follows from the first by the binomial inversion formula. \qed

The left generating functions are also related in a simple way.

55
**Theorem 6.2.** Suppose that $\Gamma$ and $\bar{\Gamma}$ denote the left generating functions for $(S,\rightarrow)$ and $(S,\bar{\rightarrow})$, with convergence functions $a$ and $\bar{a}$, respectively. Then for $x \in S$,

\[
\Gamma(x,t) = \frac{1}{1-t} \Gamma \left( x, \frac{t}{1-t} \right), \quad \bar{a}(x) = \frac{a(x)}{1 + a(x)}
\]
\[
\bar{\Gamma}(x,t) = \frac{1}{1+t} \bar{\Gamma} \left( x, \frac{t}{1+t} \right), \quad a(x) = \frac{\bar{a}(x)}{1 - \bar{a}(x)}
\]

Proof. For $x \in S$,

\[
\Gamma(x,t) = \sum_{n=0}^{\infty} \frac{\hat{\gamma}_n(x)}{\gamma_k(x)} t^n = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \left( \frac{n}{k} \right) \gamma_k(x) t^n
\]
\[
= \sum_{k=0}^{\infty} \gamma_k(x) \sum_{n=k}^{\infty} \frac{k}{n} t^n = \sum_{k=0}^{\infty} \gamma_k(x) \frac{t^k}{(1-t)^{k+1}}
\]
\[
= \frac{1}{1-t} \sum_{k=0}^{\infty} \gamma_k(x) \left( \frac{t}{1-t} \right)^k = \frac{1}{1-t} \Gamma \left( x, \frac{t}{1-t} \right)
\]

The series converges when $|t/(1-t)| \leq a(x)$. Solving $t/(1-t) = a(x)$ for $t$ gives $\bar{a}(x) = a(x)/[1 + a(x)]$. It should be understood that $\bar{a}(x) = 1$ if $a(x) = \infty$.

The second result follows from the first. For $t \in \mathbb{R}$, let $s = t/(1+t)$ so that $t = s/(1-s)$. Then

\[
\Gamma(x,t) = \Gamma \left( x, \frac{s}{1-s} \right) = (1-s)\bar{\Gamma}(x,s) = \frac{1}{1+t} \bar{\Gamma} \left( x, \frac{t}{1+t} \right)
\]

Solving $t/(1+t) = \bar{a}(x)$ for $t$ gives $a(x) = \bar{a}(x)/[1 - \bar{a}(x)]$. It should be understood that $a(x) = \infty$ if $\bar{a}(x) = 1$.

Suppose now that $X$ is a random variable with values in $S$. There are very simple relationships between the right probability functions, right rate functions, and generating functions of $X$ with respect to $(S,\rightarrow)$ and $(S,\bar{\rightarrow})$. This in turn leads to a simple relationship between the constant rate distributions for the two graphs (if they exist). As usual, we assume that $X$ is supported by the graph $(S,\rightarrow)$ (and as we will see below, supported also by $(S,\bar{\rightarrow})$).

**Theorem 6.3.** Suppose that $X$ has density function $f$.

(a) Let $F$ and $\bar{F}$ denote the right probability functions of $X$ for $(S,\rightarrow)$ and $(S,\bar{\rightarrow})$, respectively. Then

\[
\bar{F} = f + F, \quad F = \bar{F} - f
\]

(b) Let $r$ and $\bar{r}$ denote the right rate functions of $X$ for $(S,\rightarrow)$ and $(S,\bar{\rightarrow})$, respectively. Then

\[
r = r/(1+r), \quad \bar{r} = \bar{r}/(1-\bar{r})
\]

(c) Let $\Lambda$ and $\bar{\Lambda}$ denote the generating functions of $X$ for $(S,\rightarrow)$ and $(S,\bar{\rightarrow})$, respectively. Then

\[
\bar{\Lambda}(t) = \frac{1}{1-t} \Lambda \left( \frac{t}{1-t} \right)
\]
\[
\Lambda(t) = \frac{1}{1+t} \bar{\Lambda} \left( \frac{t}{1+t} \right)
\]

(d) If $X$ has constant rate $\alpha \in (0, \infty)$ for $(S,\rightarrow)$, then $X$ has constant rate $\beta = \alpha/(1+\alpha)$ for to $(S,\bar{\rightarrow})$.

(e) If $X$ has constant rate $\beta \in (0, 1)$ for $(S,\bar{\rightarrow})$, then $X$ has constant rate $\alpha = \beta/(1-\beta)$ for $(S,\rightarrow)$.

Proof. The proofs are simple.

(a) Note that

\[
\bar{F}(x) = P(x \bar{\rightarrow} X) = P(X = x) + P(x \rightarrow X) = f(x) + F(x), \quad x \in S
\]
(b) From (a),
\[ \tilde{r} = \frac{f}{F} = \frac{f}{f + F} = \frac{f/F}{f/F + 1} = \frac{r}{r+1} \]

(c) This follows from Proposition 6.2.

(d) This follows from (b).

(e) This follows from (b).

Note that \( \tilde{r} < 1 \). Also, there is a one-to-one correspondence between constant rate distributions for \((S, \rightarrow)\) and \((S, \leftarrow)\). The respective rate constants \( \alpha \) and \( \beta \) satisfy \( \frac{1}{\beta} - \frac{1}{\alpha} = 1 \), a bit reminiscent of conjugate exponents. In particular, a constant rate distribution exists for \((S, \rightarrow)\) if and only if a constant rate distribution exists for \((S, \leftarrow)\).

Suppose that \( X \) is a random variable on \( S \) as above. Let \( \bar{X} = (X_1, X_2, \ldots) \) and \( \bar{X} = (\bar{X}_1, \bar{X}_2, \ldots) \) denote the random walks on \((S, \rightarrow)\) and \((S, \leftarrow)\) associated with \( X \), respectively. The transition densities are also related in a simple way.

**Theorem 6.4.** Let \( \bar{P} \) and \( \bar{P} \) denote the transition densities of \( \bar{X} \) and \( \bar{X} \), respectively.

(a) Let \( r \) denote the right rate function of \( X \) for \((S, \rightarrow)\). Then
\[ \bar{P}(x,x) = \frac{r(x)}{r(x) + 1}, \quad x \in S; \quad \bar{P}(x,y) = \frac{1}{r(x) + 1} P(x,y), \quad x \rightarrow y \]

(b) Let \( \tilde{r} \) denote the right rate function of \( X \) for \((S, \leftarrow)\). Then
\[ P(x,y) = \frac{1}{1 - \tilde{r}(x)} \bar{P}(x,y), \quad x \rightarrow y \]

**Proof.** As before, let \( F \) and \( \bar{F} \) denote the right probability functions of \( f \) for \((S, \rightarrow)\) and \((S, \leftarrow)\), respectively. If \( x \leftarrow y \) then
\[ P(x,y) = \frac{f(y)}{\bar{F}(x)} = \frac{f(y)}{f(x) + F(x)} = \frac{f(y)/F(x)}{f(x)/F(x) + 1} = \frac{f(y)/F(x)}{r(x) + 1} \]

Part (a) follows since \( x \leftarrow y \) if and only if \( x = y \) or \( x \rightarrow y \). Part (b) follows from part (a). \( \square \)

As a corollary, the higher order-density functions for \((S, \leftarrow)\) and \((S, \rightarrow)\) have a simple relationship when the underlying random variable has constant rate.

**Corollary 6.1.** For \( n \in \mathbb{N}_+ \), let \( f_n \) and \( \tilde{f}_n \) denote the density functions of \( X_n \) and \( \bar{X}_n \), respectively.

(a) If \( X \) has constant rate \( \alpha \in (0, \infty) \) for \((S, \rightarrow)\) then
\[ \bar{f}_n = \frac{1}{(1 + \alpha)^{n-1}} \sum_{j=1}^{n} \binom{n-1}{j-1} \alpha^{n-j} f_j \]

(b) If \( X \) has constant rate \( \beta \) for \((S, \leftarrow)\) then
\[ f_n = \frac{1}{(1 - \beta)^{n-1}} \sum_{j=1}^{n} \binom{n-1}{j-1} (-\beta)^{n-j} \bar{f}_j \]

**Proof.** The proof uses the general formula for the higher-order densities in Corollary 4.12.

(a) If \( X \) has constant rate \( \alpha \in (0, \infty) \) for \((S, \rightarrow)\) then \( X \) has constant rate \( \alpha/(\alpha + 1) \) for \((S, \leftarrow)\). Hence
\[ \bar{f}_n = \left( \frac{\alpha}{\alpha + 1} \right)^n \tilde{f}_n \]

But \( \gamma_{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} \gamma_k \) and \( \bar{F} = F + F = \alpha F + F = (\alpha + 1)F \). Substituting and simplifying gives the result.
(b) The proof is similar and is left as an exercise.

Finally, recall that \( N = \{ N_A : A \in \mathcal{S} \} \) and \( \bar{N} = \{ \bar{N}_A : A \in \mathcal{S} \} \) are the point processes corresponding to the random walks \( X \) and \( \bar{X} \), respectively.

**Proposition 6.1.** Suppose that \( X \) has constant rate \( \alpha \in (0, \infty) \) for \( (S, \to) \). Then

\[
E(\bar{N}_A) = (1 + \alpha) E(N_A), \quad A \in \mathcal{S}
\]

**Proof.** Since \( X \) has constant rate \( \alpha \in (0, \infty) \) for \( (S, \to) \), \( X \) has constant rate \( \beta = \alpha / (\alpha + 1) \in (0, 1) \) for \( (S, \leftrightarrow) \). Hence using Corollary 4.13 and Proposition 6.2

\[
E(\bar{N}_A) = E[\bar{\Gamma}(X, \beta), X \in A] = E \left[ \frac{1}{1 - \beta} \Gamma \left( X, \frac{\beta}{1 - \beta} \right), X \in A \right]
\]

\[
= (1 + \alpha) E(\Gamma(X, \alpha), X \in A) = (1 + \alpha) E(N_A)
\]

**Exercise 6.1.** Consider the reflexive completion \( (S, \leftrightarrow) \) of the diamond graph \( (S, \leftrightarrow) \) introduced in Exercise 2.3.

(a) Find the walk function of order \( n \in \mathbb{N} \).

(b) Find the rate constant and the density function of the constant rate distribution.

\( \checkmark \checkmark \)
Chapter 7

Induced Graphs

A measurable function from a measure space to another measure space with a graph leads to a graph on the domain space in a natural way. This chapter explores this setting in general and in a number of important special cases.

7.1 Introduction

Suppose that \((S, \mathcal{S}, \lambda)\) and \((T, \mathcal{T}, \mu)\) are \(\sigma\)-finite measure spaces with measurable diagonals (our usual assumptions). Suppose also that \(\varphi\) is a measurable function from \(S\) onto \(T\). Then \(\varphi\) induces a measurable partition \(\mathcal{P} = \{S_t : t \in T\}\) of \(S\) where

\[ S_t = \varphi^{-1}\{t\} = \{x \in S : \varphi(x) = t\}, \quad t \in T \]

Let \(\mathcal{S}_t\) denote the associated \(\sigma\)-algebra on \(S_t\), so that

\[ \mathcal{S}_t = \{A \in \mathcal{S} : A \subseteq S_t\}, \quad t \in T \]

For \(t \in T\), the measurable space \((S_t, \mathcal{S}_t)\) also has a measurable diagonal, since

\[ \{(x, x) : x \in S_t\} = \{(x, x) : x \in S\} \cap (S_t \times S_t) \]

To complete the basic setup, we need an assumption that links the measures \(\lambda\) and \(\mu\) and the mapping \(\varphi\) in a fundamental way.

**Assumption 7.1.** We assume that there is a finite, positive measure \(\lambda_t\) on \((S_t, \mathcal{S}_t)\) for each \(t \in T\) so that

\[ t \mapsto \lambda_t(S_t) \text{ is measurable and } \lambda(A) = \int_T \lambda_t(A \cap S_t) \, d\mu(t), \quad A \in \mathcal{S} \]

It’s possible to give conditions on the measure spaces and the mapping to ensure that the basic assumption is satisfied, but such technical conditions are not worth the effort. As we will see, the basic assumption is satisfied in the important special cases that are of most interest to us—the discrete case studied in Section 7.4 of this chapter, and the norm graphs studied in Section 10.4 of Chapter 10. In the latter case, the graphs are on Euclidean spaces where the basic assumption is essentially the co-area formula. However, it’s easy to see that the assumption places significant restrictions on the two measure spaces and the mapping. Naturally, there is an integral version of the basic assumption. If \(f : S \to \mathbb{R}\) is measurable then

\[ \int_S f(x) \, d\lambda(x) = \int_T \int_{S_t} f(x) \, d\lambda_t(x) \, d\mu(t) \quad (7.1) \]

assuming of course that the integrals make sense. From now on, \(\lambda_t\) is the standard reference measure on \((S_t, \mathcal{S}_t)\) for \(t \in T\), so in particular, density functions on \(S_t\) are with respect to this measure.

Let \(\beta(t) = \lambda_t(S_t)\) for \(t \in T\), so that by assumption, \(\beta\) is a measurable function from \(T\) into \((0, \infty)\). The function \(\beta\) will play an important role, as we will see. As a first indication, recall that the measure \(\nu\) on \((T, \mathcal{T})\) induced by \(\varphi\) is given by

\[ \nu(A) = \lambda[\varphi^{-1}(A)], \quad A \in \mathcal{T} \]
The induced measure $\nu$ is typically not the same as the reference measure $\mu$, but under our basic assumption, they are closely related.

**Proposition 7.1.** The induced measure $\nu$ is absolutely continuous with respect to the reference measure $\mu$, with density function $\beta$.

*Proof.* Let $A \in \mathcal{F}$. Then by the basic assumption,

$$
\nu(A) = \lambda[\varphi^{-1}(A)] = \int_T \lambda_t[\varphi^{-1}(A) \cap S_t] d\mu(T) = \int_T \lambda_t[\varphi^{-1}(A \cap \{t\})] d\mu(t)
$$

But $\varphi^{-1}(A \cap \{t\}) = \varphi^{-1}\{t\} = S_t$ if $t \in A$ and is $\emptyset$ otherwise, so

$$
\nu(A) = \int_A \lambda_t(S_t) d\mu(t) = \int_A \beta(t) d\mu(t)
$$

Since $\beta$ is strictly positive, the reference measure $\mu$ is also absolutely continuous with respect to the induced measure $\nu$, with density function $1/\beta$. Hence, the two measures are equivalent, under the natural equivalence relation associated with absolute continuity. But again, the reference measure is usually more natural—counting measure in the discrete case and Lebesgue measure on Euclidean spaces, for example. Next is our fundamental definition.

**Definition 7.1.** Suppose that $(T, \rightarrow)$ is a graph. The graph $(S, \Rightarrow)$ induced by $(T, \rightarrow)$ and the mapping $\varphi$ is defined by $x \Rightarrow y$ if and only if $\varphi(x) \rightarrow \varphi(y)$ for $(x, y) \in S^2$.

*Proof.* We need to show that $(S, \Rightarrow)$ is a valid graph. The function $(\varphi, \varphi)$ from $S^2$ into $T^2$ is measurable, and

$$
\{(x, y) \in S^2 : x \Rightarrow y\} = \{(x, y) \in S^2 : \varphi(x) \rightarrow \varphi(y)\}
$$

is the inverse image under $(\varphi, \varphi)$ of $\{(u, v) \in T^2 : u \rightarrow v\}$.

The main point of this chapter is to see how results for the graph $(T, \rightarrow)$ can be leveraged to obtain corresponding results for the induced graph $(S, \Rightarrow)$. Going forward, we will use our usual notation for mathematical objects associated with the graphs $(S, \Rightarrow)$ and $(T, \rightarrow)$, but with the latter augmented by a circumflex.

**Proposition 7.2.** Let $\gamma_n$ denote the left walk function of order $n \in \mathbb{N}$ for the graph $(S, \Rightarrow)$. Then

$$
\gamma_n(x) = \int_{t_1 \rightarrow t_2 \rightarrow \cdots \rightarrow t_n \Rightarrow \varphi(x)} \beta(t_1)\beta(t_2)\cdots\beta(t_n) d\mu^n(t_1, t_2, \ldots, t_n)
$$

*Proof.* As before, let $\nu$ denote the measure on $(T, \mathcal{F})$ induced by $\varphi$. By the definitions, and by the change of variables theorem,

$$
\gamma_n(x) = \int_{S^n} 1(x_1 \Rightarrow x_2 \Rightarrow \cdots \Rightarrow x_n \Rightarrow x) d\lambda_n(x_1, x_2, \ldots, x_n)
$$

$$
= \int_{S^n} 1[\varphi(x_1) \rightarrow \varphi(x_2) \rightarrow \cdots \rightarrow \varphi(x_n)] \rightarrow \varphi(x_n) d\lambda_n(x_1, x_2, \ldots, x_n)
$$

$$
= \int_{T^n} 1[t_1 \rightarrow t_2 \rightarrow \cdots \rightarrow t_n \Rightarrow \varphi(x)] d\nu^n(t_1, t_2, \ldots, t_n)
$$

The result now follows from Proposition 7.1, since $d\nu(t_1, t_2, \ldots, t_n) = \beta(t_1)\beta(t_2)\cdots\beta(t_n) d\mu(t_1, t_2, \ldots, t_n)$.

In the special case that $\beta$ is constant on $T$, $\gamma_n(x) = \beta^n\hat{\gamma}_n[\varphi(x)]$ for $x \in S$, where $\hat{\gamma}_n$ is the left walk function of order $n$ for $(T, \rightarrow)$.
7.2 Distributions

Suppose now that $X$ is a random variable with values in $S$, and let $\hat{X} = \varphi(X)$. There is a trivial relationship between the right probability functions of $X$ and $\hat{X}$.

**Proposition 7.3.** Let $F$ and $\hat{F}$ denote the right probability functions of $X$ and $\hat{X}$ for the graphs $(S, \rightarrow)$ and $(T, \rightarrow)$, respectively. Then

$$F(x) = \hat{F}[\varphi(x)], \quad x \in S$$

**Proof.** By definition,

$$F(x) = \mathbb{P}(x \Rightarrow X) = \mathbb{P}[\varphi(x) \rightarrow \varphi(X)] = \mathbb{P}[\varphi(x) \rightarrow \hat{X}] = \hat{F}[\varphi(x)], \quad x \in S$$

So $F$ is constant on $S_t$ for each $t \in T$. We assume that $\hat{X}$ is supported by $(T, \rightarrow)$, so that $\hat{F}(t) > 0$ for $t \in T$. It then follows that $X$ is supported by $(S, \Rightarrow)$.

**Proposition 7.4.** Suppose that $X$ has density function $f$. Then $\hat{X}$ has density function $\hat{f}$ given by

$$\hat{f}(t) = \int_{S_t} f(x) \, d\lambda_t(x), \quad t \in T$$

**Proof.** Let $A \in \mathcal{F}$. Then

$$\mathbb{P}(\hat{X} \in A) = \mathbb{P}[X \in \varphi^{-1}(A)] = \int_{\varphi^{-1}(A)} f(x) \, d\lambda(x) = \int_S 1[x \in \varphi^{-1}(A)] f(x) \, d\lambda(x)$$

So by (B),

$$\mathbb{P}(\hat{X} \in A) = \int_T \int_{S_t} 1[x \in \varphi^{-1}(A)] f(x) \, d\lambda_t(x) \, d\mu(t) = \int_T \int_S 1[x \in \varphi^{-1}(A) \cap S_t] f(x) \, d\lambda_t(x) \, d\mu(t)$$

But $\varphi^{-1}A \cap S_t = \varphi^{-1}(A) \cap \varphi^{-1}\{t\} = \varphi^{-1}(A \cap \{t\})$ and this set is $S_t$ if $t \in A$ and is $\emptyset$ otherwise. Hence

$$\mathbb{P}(\hat{X} \in A) = \int_A \int_{S_t} f(x) \, d\lambda_t(x) \, d\mu(t)$$

**Corollary 7.1.** For $t \in T$, a conditional density of $X$ given $\hat{X} = t$ is defined by

$$f(x \mid t) = \frac{f(x)}{\hat{f}(t)}, \quad x \in S_t$$

**Proof.** Fix $t \in T$. For $A \in \mathcal{F}$,

$$\int_T \hat{f}(t) \int_{A \cap S_t} f(x \mid t) \, d\lambda_t(x) \, d\mu(t) = \int_T \hat{f}(t) \int_{A \cap S_t} \frac{f(x)}{\hat{f}(t)} \, d\lambda_t(x) \, d\mu(t)$$

$$= \int_T \int_{A \cap S_t} f(x) \, d\lambda_t(x) \, d\mu(t) = \int_A f(x) \, d\lambda(x) = \mathbb{P}(X \in A)$$

where again we have used (B). So it follows by definition that

$$\mathbb{P}(X \in A \mid \hat{X} = t) = \int_A f(x \mid t) \, d\lambda_t(x), \quad A \in \mathcal{F}_t$$

**Corollary 7.2.** Let $r$ and $\hat{r}$ denote the rate functions of $X$ and $\hat{X}$ for the graphs $(S, \Rightarrow)$ and $(T, \rightarrow)$, respectively. Then

$$r(x) = f[x \mid \varphi(x)]\hat{r}[\varphi(x)], \quad x \in S$$
The conditional distribution of $X_L$ Conversely, suppose that (a) and (b) hold, so in particular a density $\hat{f}$ probability functions of $X$ for $(S, \Rightarrow)$ and $(T, \rightarrow)$, respectively. Then

$$r(x) = \frac{f(x)}{F(x)} = \frac{f(x)}{f[\varphi(x)]} = f[x \mid \varphi(x)]\hat{r} \varphi(x), \quad x \in S$$

So $r(x) = f(x \mid t)\hat{r}(t)$ for $x \in S_t$ and $t \in T$, and hence

$$\int_{S_t} r(x) \, d\lambda_t(x) = \hat{r}(t), \quad t \in T$$

As usual, we are particularly interested in constant rate distributions for the induced graph $(S, \Rightarrow)$. Here is the main result.

**Theorem 7.1.** $X$ has constant rate $\alpha \in (0, \infty)$ for $(S, \Rightarrow)$ if and only if

(a) $\hat{X}$ has rate function $\hat{r}$ for $(T, \rightarrow)$ given by $r(t) = \alpha \beta(t)$ for $t \in T$.

(b) The conditional distribution of $X$ given $\hat{X} = t$ is uniform on $S_t$ for $t \in T$.

**Proof.** As before, let $\hat{F}$ denote the right probability function of $\hat{X}$ for $(T, \rightarrow)$, so that the right probability function $F$ of $X$ for $(S, \Rightarrow)$ is given by $F(x) = \hat{F}[\varphi(x)]$ for $x \in S$. Suppose first that $X$ has constant rate $\alpha \in (0, \infty)$ for $(S, \Rightarrow)$. Then $f = \alpha F$ is a density of $X$ and hence by Proposition 7.4, a density $\hat{f}$ for $\hat{X}$ is given by

$$\hat{f}(t) = \int_{S_t} f(x) \, d\lambda_t(x) = \int_{S_t} \alpha \hat{F}[\varphi(x)] \, d\lambda_t(x) = \alpha \hat{F}(t) \beta(t), \quad t \in T$$

Hence the rate function of $\hat{X}$ for $(T, \rightarrow)$ is $\hat{r} = \alpha \beta$. Moreover, by Corollary 7.1, for $t \in T$, the conditional density of $X$ given $\hat{X} = t$ is

$$f(x \mid t) = \frac{f(x)}{f(t)} = \frac{\alpha F(x)}{\alpha F(t)} = \frac{\hat{F}(t)}{\beta(t)} = \frac{1}{\beta(t)}, \quad x \in S_t$$

Conversely, suppose that (a) and (b) hold, so in particular a density $\hat{f}$ of $\hat{X}$ is given by $\hat{f}(t) = \alpha \beta(t)\hat{F}(t)$ for $t \in T$. Let $f = \alpha F$ so that $f(x) = \alpha \hat{F}([\varphi(x)])$ for $x \in S$. We need to show that $f$ is a density of $X$. For $A \in \mathcal{S}$,

$$\mathbb{P}(X \in A) = \mathbb{E}[\mathbb{P}(X \in A \mid \hat{X})] = \int_T \hat{f}(t) \mathbb{P}(X \in A \mid \hat{X} = t) \, d\mu(t)$$

$$= \int_T \hat{f}(t) \frac{\lambda_t(A \cap S_t)}{\lambda_t(S_t)} \, d\mu(t) = \int_T \alpha \beta(t)\hat{F}(t) \frac{\lambda_t(A \cap S_t)}{\beta(t)} \, d\mu(t)$$

$$= \int_T \alpha \hat{F}(t) \lambda_t(A \cap S_t) \, d\mu(t)$$

But on the other hand, $f(x) = \alpha \hat{F}(t)$ for $x \in S_t$ and $t \in T$, so by (B),

$$\int_A f(x) \, d\lambda(x) = \int_T \int_{A \cap S_t} f(x) \, d\lambda_t(x) \, d\mu(t) = \int_T \alpha \hat{F}(t) \lambda_t(A \cap S_t)$$

Let $B$ be the kernel for $(T, \rightarrow)$ defined by

$$B(u, v) = \beta(u)1(u \rightarrow v) \quad (u, v) \in T^2$$

Then condition (a) in Theorem 7.1 means that $\alpha$ is a right eigenvalue of $B$ and $\hat{f}$ is a corresponding eigenfunction, for the space $\mathcal{L}_1(T)$. Note also that if $\beta$ is constant on $T$, then $X$ has constant rate $\alpha$ for the graph $(S, \Rightarrow)$ if and only if $\hat{X}$ has constant rate $\alpha \beta$ for the graph $(T, \rightarrow)$.
**Proposition 7.5.** Let $H$ denote the entropy operator, and suppose again that $X$ is a random variable on $S$. Then

$$H(X) = H(\hat{X}) + H(X | \hat{X})$$

**Proof.** This is a general result in entropy, since the distribution of $X$ completely determines the distribution of $(X, X)$, but we give a separate proof in this context.

$$H(X) = -E[\ln f(X)] = -E[E[\ln f(X) | \hat{X}]] = -E(E[\ln(\hat{f}(X)f(X \mid \hat{X})) | \hat{X}])$$

$$= -E[\ln \hat{f}(X)] - E(E[\ln f(X \mid \hat{X}) | \hat{X}]) = H(\hat{X}) + H(X | \hat{X})$$

\[
\]

### 7.3 Random Walks

Let $X$ be a random variable on $S$ with probability density function $f$ and with right probability function $F$ for $(S, \Rightarrow)$. Let $\hat{X} = \varphi(X)$ be the corresponding variable on $T$ with probability density function $\hat{f}$ and with right probability function $\hat{F}$ for $(T, \rightarrow)$. As before, we assume that $\hat{f}(t) > 0$ for almost all $t \in T$.

**Theorem 7.2.** Suppose that $X = (X_1, X_2, \ldots)$ is the random walk on $(S, \Rightarrow)$ associated with $X$. Then $\hat{X} = (\hat{X}_1, \hat{X}_2, \ldots)$ is the random walk on $(T, \rightarrow)$ associated with $\hat{X}$. Let $P$ and $\hat{P}$ denote the transition densities for $X$ and $\hat{X}$, respectively. Then for $n \in \mathbb{N}_+$,

$$P^n(x, y) = \hat{P}^n[\varphi(x), \varphi(y)]f[y \mid \varphi(y)], \quad (x, y) \in S^2$$

**Proof.** By definition, $X_1$ has density function $f$ so $\hat{X}_1$ has density function $\hat{f}$. Moreover, $X$ is a homogenous, discrete-time Markov process with transition density $P$ given by $P(x, y) = f(y)/F(x)$ if $x \Rightarrow y$ (and 0 otherwise). That is,

$$P(x, y) = \frac{f(y)}{F(x)} = \frac{f(y)}{\frac{\hat{f}(\varphi(y))}{\hat{F}([\varphi(y)])}} = f(y \mid \varphi(y))\hat{P}(\varphi(x), \varphi(y))$$

The important point is that $P(x, y)$ is constant for $x \in S_t$ and $t \in T$, so for $k \in \mathbb{N}_+ \cup \{0\}$ conditioning on $X_k = x$ is the same as conditioning on $X_k \in T$. Let $B \in \mathcal{F}$, $n \in \mathbb{N}_+$, and $(t_1, t_2, \ldots, t_n) \in T^n$. Then

$$P(\hat{X}_{n+1} \in B \mid \hat{X}_1 = t_1, \ldots, \hat{X}_{n} = t_n) = \frac{1}{F(t)}\int_B \int_{S_n} f(x) d\lambda_n(x) d\mu(u) = \frac{1}{F(t)}\int_B \hat{f}(u) d\mu(u) = \int_B \hat{P}(t, u) d\mu(u)$$

For the higher-order transition densities, we need the relation between the higher-order kernels. For $n \in \mathbb{N}_+$,

$$R_n(x, y) = \int_{x \Rightarrow x_1, \ldots, x_n \Rightarrow y} r(x_1) \cdots r(x_n) d\lambda^n(x_1, \ldots, x_n)$$

$$= \int_{\varphi(x) \Rightarrow \varphi(x_1), \ldots, \varphi(x_n) \Rightarrow \varphi(y)} r(x_1) \cdots r(x_n) d\lambda^n(x_1, \ldots, x_n)$$

$$= \int_{\varphi(x) \Rightarrow t_1, \ldots, t_n \Rightarrow \varphi(y)} \int_{S_{t_1} \times \cdots \times S_{t_n}} r(x_1) \cdots r(x_n) d(\lambda_{t_1} \cdots \lambda_{t_n})(x_1, \ldots, x_n) d\mu^n(t_1, \ldots, t_n)$$

$$= \int_{\varphi(x) \Rightarrow t_1, \ldots, t_n \Rightarrow \varphi(y)} \hat{r}(t_1) \cdots \hat{r}(t_n) d\mu^n(t_1, \ldots, t_n) = \hat{R}_n[\varphi(x), \varphi(y)], \quad (x, y) \in S^2$$

Hence

$$P^n(x, y) = R_n(x, y)P(x, y) = \hat{R}_n[\varphi(x), \varphi(y)]\hat{P}[\varphi(x), \varphi(y)]f[y \mid \varphi(y)]$$

$$= \hat{P}^n[\varphi(x), \varphi(y)]\hat{f}[y \mid \varphi(y)], \quad (x, y) \in S^2$$

\[
\]

\[
\]
Note in particular that
\[ \int_{S_t} P^n(x, y) \, d\lambda_t(y) = \tilde{P}^n[\varphi(x), t], \quad x \in S, \; t \in T \]
The next two corollaries give more information about the connections between the random walks \( X \) and \( \tilde{X} \).

**Corollary 7.3.** Consider again the random walks \( X \) and \( \tilde{X} \) and fix \( n \in \mathbb{N}_+ \).

(a) Let \( g_n \) and \( \tilde{g}_n \) denote the density functions of \( (X_1, X_2, \ldots, X_n) \) and \( (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n) \), respectively. For \((x_1, x_2, \ldots, x_n) \in S^n\),
\[ g_n(x_1, x_2, \ldots, x_n) = f(x_1 | \varphi(x_1)) f([x_2 | \varphi(x_2)] \cdots f(x_n | \varphi(x_n)) \tilde{g}_n[\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_n)] \]
(b) Let \( f_n \) and \( \tilde{f}_n \) denote the density functions of \( X_n \) and \( \tilde{X}_n \), respectively. Then
\[ f_n(x) = f(x | \varphi(x)) \tilde{f}_n[\varphi(x)], \quad x \in S \]

**Proof.** The results follow easily from the basic theory.

(a) Recall that
\[ g_n(x_1, x_2, \ldots, x_n) = r(x_1)r(x_2) \cdots r(x_{n-1})f(x_n), \quad (x_1, x_2, \ldots, x_n) \in S^n \]
The result then follows since \( r(x) = f(x | \varphi(x)) \tilde{f}[\varphi(x)] \) and \( f(x) = f(x | \varphi(x)) \tilde{f}[\varphi(x)] \) for \( x \in S \).
(b) We need the relation between the order \( n \) functions for \( n \in \mathbb{N}_+ \). The same argument used to show that \( R_n(x, y) = R_n[\varphi(x), \varphi(y)] \) for \((x, y) \in S^2 \) shows that \( v_n(x) = \tilde{v}_n[\varphi(x)] \) for \( x \in S \). So now
\[ f_n(x) = v_{n-1}(x) f(x) = \tilde{v}_{n-1}[\varphi(x)] f(x | \varphi(x)) \tilde{f}[\varphi(x)] = f(x | \varphi(x)) \tilde{f}_n[\varphi(x)], \quad x \in S \]

In particular,
\[ \int_{S_{t_1} \times \cdots \times S_{t_n}} g_n(x_1, \ldots, x_n) \, d(\lambda_{t_1} \cdots \lambda_{t_n})(x_1, \ldots, x_n) = \tilde{g}(t_1, \ldots, t_n), \quad (t_1, \ldots, t_n) \in T^n \]
\[ \int_{S_t} f_n(x) \, d\lambda_t(x) = \tilde{f}_n(t), \quad t \in T \]

**Corollary 7.4.** Consider again the random walks \( X = (X_1, X_2, \ldots) \) and \( \tilde{X} = (\tilde{X}_1, \tilde{X}_2, \ldots) \).

(a) \( X \) is a conditionally independent sequence given \( \tilde{X} \).
(b) For \( n \in \mathbb{N}_+ \) and \( t \in T \), the conditional distribution of \( X_n \) given \( \tilde{X}_n = t \) has density function \( f(\cdot | t) \).

**Proof.** The results follow directly from Corollary ??.

(a) Let \( n \in \mathbb{N}_+ \) and \((t_1, t_2, \ldots, t_n) \in T^n \). Then the conditional density of \((X_1, X_2, \ldots, X_n) \) given \( \{\tilde{X}_1 = t_1, \tilde{X}_2 = t_2, \ldots, \tilde{X}_n = t_n\} \) is
\[ (x_1, x_2, \ldots, x_n) \mapsto \frac{g_n(x_1, x_2, \ldots, x_n)}{\tilde{g}_n(t_1, \ldots, t_n)} = f(x_1 | t_1) f(x_2 | t_2) \cdots f(x_n | t_n), \quad (x_1, x_2, \ldots, x_n) \in S_{t_1} \times S_{t_2} \times \cdots \times S_{t_n} \]
So by the basic factorization theorem, it follows that \((X_1, X_2, \ldots, X_n) \) are conditionally independent given \( \{\tilde{X}_1 = t_1, \tilde{X}_2 = t_2, \ldots, \tilde{X}_n = t_n\} \) and that \( X_k \) has conditional density function \( f(\cdot | t_k) \).
(b) This does not follow immediately from (a) since we are conditioning only on \( \tilde{X}_n \), but the proof is just as easy. For \( t \in T \), the conditional density of \( X_n \) given \( \tilde{X}_n = t \) is
\[ x \mapsto \frac{f_n(x)}{f_n(t)} = f(x | t), \quad x \in S_t \]
The following theorem gives the connection for the point processes $N$ and $\hat{N}$ associated with the random walks $X$ and $\hat{X}$.

**Theorem 7.3.** Define the measure $\eta$ on $(T, \mathcal{F})$ by $\eta(B) = E(\hat{N}_B)$ for $B \in \mathcal{F}$. Then

$$\mathbb{E}(N_A) = \int_T \mathbb{P}(X \in A \cap S_t \mid \hat{X} = t) \, d\eta(t), \quad A \in \mathcal{F}$$

**Proof.** Note that

$$\mathbb{E}(N_A) = \sum_{n=0}^{\infty} \mathbb{P}(X_n \in A) = \sum_{n=0}^{\infty} \int_A f_n(x) \, d\lambda(x)$$

$$= \sum_{n=0}^{\infty} \int_T \int_{A \cap S_t} f_n(x) \, d\lambda_t(x) \, d\mu(t) = \sum_{n=0}^{\infty} \int_T \int_{A \cap S_t} f(x \mid t) \hat{f}_n(t) \, d\lambda_t(x) \, d\mu(t)$$

$$= \sum_{n=0}^{\infty} \int_T \hat{f}_n(t) \int_{A \cap S_t} f(x \mid t) \, d\mu(t) = \sum_{n=0}^{\infty} \int_T \hat{f}_n(t) \mathbb{P}(X \in A \cap S_t \mid \hat{X} = t) \, d\mu(t)$$

$$= \int_T \mathbb{P}(X \in A \cap S_t \mid \hat{X} = t) \, d\eta(t)$$

since $t \mapsto \sum_{n=0}^{\infty} \hat{f}_n(t)$ is the density of $\eta$ with respect to $\mu$. \qed

In the case that $X$ has constant rate $\alpha \in (0, \infty)$ for the graph $(S, \Rightarrow)$, the random walk results simplify significantly since $f(x \mid i) = 1 / \beta(t)$ for $x \in S_t$, the density function of the uniform distribution on $S_t$ for $t \in T$. Here is a summary of the results, using the same notation as above.

**Corollary 7.5.** Suppose that $X$ has constant rate $\alpha \in (0, \infty)$, and that $X = (X_1, X_2, \ldots)$ and $\hat{X} = (\hat{X}_1, \hat{X}_2, \ldots)$ are the random walks on $(S, \Rightarrow)$ and $(I, \rightarrow)$ corresponding to $X$ and $\hat{X}$ respectively.

(a) For $n \in \mathbb{N}_+$, $P^n(x, y) = \hat{P}^n[\varphi(x), \varphi(y)] / \beta[\varphi(y)]$ for $(x, y) \in S^2$.

(b) For $n \in \mathbb{N}_+$

$$g_n(x_1, x_2, \ldots, x_n) = \frac{\hat{g}_n[\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_n)]}{\beta[\varphi(x_1)] \beta[\varphi(x_2)] \cdots \beta[\varphi(x_n)]}, \quad (x_1, x_2, \ldots, x_n) \in S^n$$

(c) For $n \in \mathbb{N}_+$, $f_n(x) = \hat{f}_n[\varphi(x)] / \beta[\varphi(x)]$ for $x \in S$.

(d) Given $\hat{X}, X$ is a sequence of independent variables, and given $\hat{X}_n = t, X_n$ is uniformly distributed on $S_t$ for each $n \in \mathbb{N}_+$ and $t \in I$.

### 7.4 The Discrete Case

Suppose that the measure space $(S, \mathcal{F}, \lambda)$ is general, as above, but that the second measure space $(I, \mathcal{I}, \#)$ is discrete, so that $I$ is countable, $\mathcal{I}$ is the collection of all subsets of $I$, and the reference measure $\#$ is counting measure. Once again, $\varphi$ is a measurable function from $S$ onto $I$, so that $\mathcal{P} = \{S_i : i \in I\}$ is a countable, measurable partition of $S$. We assume that $\beta(i) = \lambda(S_i) \in (0, \infty)$ for $i \in I$, and this in turn guarantees Assumption 7.1 with $\lambda_i$ simply $\lambda$ restricted to $(S_i, \mathcal{I}_i)$ for $i \in I$. That is,

$$\lambda(A) = \sum_{i \in I} \lambda(A \cap S_i) = \sum_{i \in I} \lambda_i(A \cap S_i), \quad A \in \mathcal{F}$$

To complete the setup, $(I, \rightarrow)$ is a discrete graph and $(S, \Rightarrow)$ the corresponding induced graph. The general results in the first three sections simplify, with integrals over $T$ replaced by sums over $I$. In particular, the left walk function $\gamma_n$ of order $n \in \mathbb{N}_+$ is given by

$$\gamma_n(x) = \sum_{i_1 \rightarrow i_2 \rightarrow \cdots \rightarrow i_n \rightarrow \varphi(x)} \beta(i_1) \beta(i_2) \cdots \beta(i_n), \quad x \in S$$
Suppose now that \( X \) is a random variable with values in \( S \) and density function \( f \). The corresponding index variable \( \hat{X} = \varphi(X) \) takes values in \( I \) and has density function \( \hat{f} \) given by
\[
\hat{f}(i) = \mathbb{P}(\hat{X} = i) = \mathbb{P}(X \in S_i) = \int_{S_i} f(x) d\lambda(x), \quad t \in I
\]

**Theorem 7.4.** Suppose that \((I, \rightarrow)\) is a finite, strongly connected graph. Then there exists a unique constant rate distribution for the induced graph \((S, \Rightarrow)\).

**Proof.** As with Theorem 4.10, this is just the Perron-Frobenius theorem. The matrix \( B \) defined in (7.2) has a unique positive eigenvalue with multiplicity 1 (the largest eigenvalue) and a corresponding eigenfunction that is strictly positive. The normalized eigenfunction \( \hat{f} \) is a probability density function on \( I \) that satisfies part (a) in Theorem 7.1. Then \( f \) defined by \( f(x) = \hat{f}(\varphi(x))/\beta(\varphi(x)) \) for \( x \in S \) is the probability density function of a constant rate distribution for \((S, \Rightarrow)\). \(\square\)

Suppose that \( A \in \mathcal{S} \). Then
\[
\mathbb{E}(N_A) = \sum_{i \in I} \mathbb{P}(X \in A \cap A_i \mid X \in S_i) \mathbb{E}(\hat{N}_i)
\]

### 7.4.1 Equivalence Relations

Equivalence relations provide one of the simplest examples of induced graphs. The discrete graph is \((I, =)\), so that the only edges are loops at each vertex. Hence the corresponding relation \( \equiv \) on \( S \) is given by \( x \equiv y \) if and only if \( x, y \in S_i \) for some \( i \in I \). So as the notation suggests, \( \equiv \) is the equivalence relation associated with the partition \( \mathcal{P} \) (or equivalently the index function \( \varphi \)). The sets in the partition are the equivalence classes and the relation \( \equiv \) is reflexive, symmetric, and transitive. By symmetry, all left and right objects are the same, so we can drop the adjectives. Except for the special symbols for the relations, we use the definitions, notation, and assumptions in the first three sections.

Many of the general results simplify considerably. Of course \( S = \bigcup_{i \in I} S_i \) but also \( \equiv \) is the set \( \bigcup_{i \in I} S_i^2 \), and more generally the set of walks of length \( n \in \mathbb{N}_+ \) is the set \( \bigcup_{i \in I} S_i^{n+1} \). For \( n \in \mathbb{N} \), the walk function \( \gamma_n \) for \((S, \equiv)\) is given by
\[
\gamma_n(x) = \beta^n(i), \quad x \in S_i, \quad i \in T
\]
So \( \gamma_n = \gamma^n \) for \( n \in \mathbb{N} \). The generating function \( \Gamma \) of \((S, \equiv)\) is given by
\[
\Gamma(x, t) = \frac{1}{1 - \beta(i)t}, \quad i \in I, \quad x \in S_i, \quad |t| < 1/\beta(i)
\]
Let \( A \) denote the adjacency kernel of \((S, \equiv)\). If \( n \in \mathbb{N}_+ \), and \( u \in \mathcal{M}_+(S) \) or \( u \in \mathcal{L}_1(S) \) then
\[
A^n u(x) = \beta^{n-1}(i) \int_{S_i} u(y) d\lambda(y), \quad i \in I, \quad x \in S_i
\]

Suppose now that \( X \) is a random variable with values in \( S \), and as usual, let \( \hat{X} \) denote the corresponding index variable with values in \( I \). For the probability functions of the graphs, note that \( \hat{F} = \hat{f} \) and hence \( F(x) = \hat{f}(i) \) for \( i \in I \) and \( x \in S_i \). For the rate functions of the graphs, note that \( \bar{r}(i) = 1 \) for \( i \in I \) and \( r(x) = f(x)/\hat{f}(i) = f(x \mid i) \) for \( i \in I \) and \( x \in S_i \). So every distribution on \( I \) has constant rate 1 for \((I, =)\). Constant rate distributions for \((S, \equiv)\) exist if and only if the equivalence classes have the same size.

**Theorem 7.5.** A constant rate distribution for \((S, \equiv)\) exists if and only if \( \beta \) is constant on \( I \). In this case, the rate constant is \( \alpha = 1/\beta \) and the constant rate distribution is uniform on \( S_i \) for each \( i \in I \).

**Proof.** This follows directly from Theorem 7.1 since condition (a) becomes \( 1 = \alpha \beta(i) \) for \( i \in I \). \(\square\)

So if \( \beta \) is constant on \( I \) then there is an infinite family of distributions with constant rate \( \alpha = 1/\beta \), parametrized by \( \hat{f} \). That is, given a density function \( \hat{f} \) on \( I \), the function \( f \) on \( S \) defined by \( f(x) = \hat{f}(i)/\beta \) for \( x \in S_i \) and \( i \in I \) is a density function that has constant rate \( 1/\beta \) for \((S, \equiv)\). Random walks on \((S, \equiv)\) are particularly simple. Once again, suppose that \( X \) is a random variable on \( S \) and that \( \hat{X} \) is the corresponding index variable.
Proposition 7.6. Suppose that $X = (X_1, X_2, \ldots)$ is the random walk on $(S, \Rightarrow)$ associated with $X$, and that $N = \{N_A : A \in \mathcal{F}\}$ is the corresponding point process. Then

(a) The transition density $P$ of $X$ is given by $P(x,y) = f(y | i)$ for $i \in I$ and $x, y \in S_i$ and more generally $P^n = P^n$ for $n \in \mathbb{N}_+$.

(b) For $n \in \mathbb{N}_+$, $(X_1, X_2, \ldots, X_n)$ has density function $g_n$ given by

$$g_n(x_1, x_2, \ldots, x_n) = f(x_1) f(x_2 | i) \cdots f(x_n | i), \quad i \in I, \quad (x_1, x_2, \ldots, x_n) \in S^n_i$$

(c) For $n \in \mathbb{N}_+$, $X_n$ has density function $f_n = f$ for $n \in \mathbb{N}_+$.

(d) Given $X_1 \in S_i$ for $i \in I$, $X$ is a sequence of independent variables with common density function $f(\cdot | i)$.

(e) For $A \in \mathcal{F}$, $E(N_A) = \infty$ if $\lambda(A) > 0$ and $E(N_A) = 0$ if $\lambda(A) = 0$.

Proof. The results follow from Theorem ?? and its corollaries since the corresponding random walk $\hat{X} = (\hat{X}_1, \hat{X}_2, \ldots)$ on $(I, =)$ is constant: $\hat{X}_n = \hat{X}_1$ for $n \in \mathbb{N}_+$.

Suppose now that $X = (X_1, X_2, \ldots)$ is the random walk on $(S, \equiv)$ associated with a random variable $X$ that has constant rate $\alpha \in (0, \infty)$, so that $\beta(i) = 1/\alpha$ for $i \in I$. The density function $f$ of $X$ is constant on the equivalence classes, and $f(x) = \alpha f(i)$ for $x \in S_i$ and $i \in I$. The transition density $P$ simplifies to $P(x,y) = \alpha$ for $x, y \in S_i$ and $i \in I$. For $n \in \mathbb{N}_+$, the density function $g_n$ of $(X_1, X_2, \ldots, X_n)$ simplifies to $g_n(x_1, x_2, \ldots, x_n) = \alpha^n f(i)$ for $(x_1, x_2, \ldots, x_n) \in S^n_i$ and $i \in I$. This is a mixture of uniform distributions on $S^n_i$ for $i \in I$, with mixture density $\hat{f}$. Equivalently, given $X_1 \in S_i$, the variables $X$ is a sequence of independent variables, each uniformly distributed on $S_i$.

The following exercises explore some very simple special cases:

Exercise 7.1 (The Complete Relation). Consider the general measure space $(S, \mathcal{F}, \lambda)$ where the only set in the partition is $S$ itself, so that $\equiv$ is the set $S^2$ and is the complete relation. This corresponds to the discrete graph $(I, =)$ where $I$ is a singleton. Let $\beta = \lambda(S) \in (0, \infty)$.

(a) Give the walk function $\gamma_n$ of order $n \in \mathbb{N}_+$ in closed form.

(b) Give the generating function $\Gamma$ in closed form.

(c) Identify the constant rate distribution.

(d) Characterize the random walk on $(S, \equiv)$ associated with a distribution on $S$.

Exercise 7.2 (Equality). At the opposite extreme, suppose that the underlying measure space $(S, \mathcal{F}, \lambda)$ is discrete and that the equivalence relation is simply equality $\equiv$, so that the graphs $(S, =)$ and $(I, =)$ are essentially the same.

(a) Give the walk function $\gamma_n$ of order $n \in \mathbb{N}_+$ in closed form.

(b) Give the generating function $\Gamma$ in closed form.

(c) Identify the constant rate distribution.

(d) Characterize the random walk on $(S, =)$ associated with a distribution on $S$.

Exercise 7.3. Suppose that $S = [0, \infty)$ with the usual Borel $\sigma$-algebra $\mathcal{F}$ and Lebesgue measure $\lambda$. Let $S_k = [k, k+1)$ for $k \in \mathbb{N}$.

(a) Give the walk function $\gamma_n$ of order $n \in \mathbb{N}_+$ in closed form.

(b) Give the generating function $\Gamma$ in closed form.

(c) Identify the constant rate distribution.
7.4.2 Complete Multipartite Graphs

Complete multipartite graphs are another simple example of induced graphs. In this case, the discrete graph is \((I, \neq)\), the complete graph on \(I\) (without loops) and the complement of the graph used in previous section for equivalence relations. Thus the relation \(\leftrightarrow\) on \(S\) defined by \((I, \neq)\) is given by \(x \leftrightarrow y\) if and only if \(x \in S_i\) and \(y \in S_j\) for distinct \(i, j \in I\).

**Definition 7.2.** The graph \((S, \leftrightarrow)\) is the complete multipartite graph associated with the partition \(\mathcal{P}\). In the special case that \(#(I) = 2\), \((S, \leftrightarrow)\) is a complete bipartite graph and in the case that \(#(I) = 3\), \((S, \leftrightarrow)\) is a complete tripartite graph.

The relation \(\leftrightarrow\) is symmetric and anti-reflexive. In particular, all left and right objects are the same and so we can drop the adjectives. Except for the special symbols used for the relations, we will use the same definitions, notation, and assumptions as in the first three sections.

For \(n \in \mathbb{N}_+\) note that a walk \((i_1, i_2, \ldots, i_{n+1})\) of length \(n\) in \((I, \neq)\) is just a sequence in \(I^n\) with \(i_{k+1} \neq i_k\) for \(k \in \{1, 2, \ldots, n\}\). So from the general results in Section 1, the walk function of order \(n\) is given by

\[
\gamma_n(x) = \sum_{i_1 \neq i_2 \neq \cdots \neq i_n \neq i} \beta(i_1)\beta(i_2)\cdots\beta(i_n), \quad i \in I, \; x \in S_i
\]

In particular,

\[
\gamma(x) = \sum_{j \neq i} \beta(j) = \lambda(S) - \beta(i) \quad i \in I, \; x \in S_i
\]

Since most of our theory depends on local finiteness of the graph, we will assume from now on that \(\lambda(S) = \sum_{i \in I} \beta(i) < \infty\), so that \((S, \mathcal{P}, \lambda)\) is a finite measure space. Of course, this hold automatically if \(I\) is finite.

The walk function and generating function have simple, closed forms only in some special cases.

Suppose now that \(X\) is a random variable with values in \(S\). As usual, define the index variable \(\hat{X} = \varphi(X)\) with values in \(I\) and with density function \(\hat{f}\) given by

\[
\hat{f}(i) = \mathbb{P}(\hat{X} = i) = \mathbb{P}(X \in S_i), \quad i \in I
\]

The probability function \(\hat{F}\) of \(\hat{X}\) for \((I, \neq)\) is given by

\[
\hat{F}(i) = \sum_{j \neq i} \hat{f}(j) = 1 - \hat{f}(i), \quad i \in I
\]

and so the rate function \(\hat{r}\) of \(\hat{X}\) for \((I, \neq)\) is given by

\[
\hat{r}(i) = \frac{\hat{f}(i)}{1 - \hat{f}(i)}, \quad i \in I
\]

Note that \(\hat{r}(i)\) is the odds ratio of the event \(\{X \in S_i\}\) for \(i \in I\). It follows that the probability function \(F\) of \(X\) for \((S, \leftrightarrow)\) is given by \(F(x) = 1 - \hat{f}(i)\) for \(i \in I\) and \(x \in S_i\). The basic equation in part (a) of Theorem 7.6, for a distribution on \((S, \leftrightarrow)\) with constant rate \(\alpha \in (0, \infty)\) is \(\hat{f}(i) = \alpha\beta(i)[1 - \hat{f}(i)]\) for \(i \in I\), or equivalently,

\[
\hat{f}(i) = \frac{\alpha\beta(i)}{1 + \alpha\beta(i)}, \quad i \in I
\]

Here is our main result.

**Theorem 7.6.** There exists a unique constant rate distribution for the complete multipartite graph \((S, \leftrightarrow)\). The rate is the unique solution \(\alpha \in (0, \infty)\) of the equation

\[
\sum_{i \in I} \frac{\alpha\beta(i)}{1 + \alpha\beta(i)} = 1
\]

and the density function of the constant rate distribution is given by

\[
f(x) = \frac{\alpha}{1 + \alpha\beta(i)}, \quad i \in I, \; x \in S_i
\]
Proof. The equation for the rate constant and the form of \( f \) follow from Theorem 7.1. So we just need to show that the equation has a unique solution. Define \( \psi \) on \([0, \infty)\) by \( \psi(\alpha) = \sum_{i \in I} \frac{\alpha \beta(i)}{1 + \alpha \beta(i)} \). First we note that \( \psi(\alpha) < \infty \) for all \( \alpha \in (0, \infty) \). This is obvious of course if \( I \) is finite. In the case that \( I \) is countably infinite, we can take \( I = \mathbb{N}_+ \). Since \( \sum_{i=1}^{\infty} \beta(i) < \infty \), it follows that \( \beta(i) \to 0 \) as \( i \to \infty \). Comparing the series \( \sum_{i=1}^{\infty} \alpha \beta(i)/[1 + \alpha \beta(i)] \) with the convergent series \( \sum_{i=1}^{\infty} \beta(i) \) we have

\[
\frac{\alpha \beta(i)/[1 + \alpha \beta(i)]}{\beta(i)} = \frac{\alpha}{1 + \alpha \beta(i)} \to \alpha \quad \text{as} \quad i \to \infty
\]

Hence \( \psi(\alpha) = \sum_{i=1}^{\infty} \alpha \beta(i)/[1 + \alpha \beta(i)] < \infty \) for \( \alpha \in [0, \infty) \). Next, since \( \alpha \mapsto \alpha \beta(i)/[1 + \alpha \beta(i)] \) is increasing on \([0, \infty)\) for each \( i \), the function \( \psi \) is also increasing on \([0, \infty)\). Moreover, \( \psi(0) = 0 \) and by the monotone convergence theorem,

\[
\lim_{\alpha \to \infty} \psi(\alpha) = \sum_{i \in I} \lim_{\alpha \to \infty} \frac{\alpha \beta(i)}{1 + \alpha \beta(i)} = \sum_{i \in I} 1 = \#(I) > 1
\]

Also by the monotone convergence theorem, the function \( \psi \) is continuous. By the intermediate value theorem, there exists a unique \( \alpha \in (0, \infty) \) with \( \psi(\alpha) = 1 \).

Explicit results are difficult except in a few special cases. The case where the partition sets have the same size is particularly simple and is left as an exercise

**Exercise 7.4.** Suppose that \( \#(I) = k \in \{2, 3, \ldots\} \) and \( \beta_i = \beta \in (0, \infty) \) for \( i \in I \).

(a) Give the walk function \( \gamma_n \) of order \( n \in \mathbb{N}_+ \) in closed form.

(b) Give the generating function \( \Gamma \) in closed from.

(c) Identify the constant rate distribution.

Another tractable case is when \( \#(I) = 2 \) so that \( (I, \neq) \) is simply the (undirected) path on two vertices. For the remainder of this section, suppose \( I = \{0, 1\} \), so the partition sets are \( S_0 \) and \( S_1 = S_0^{-1} \), with sizes \( \beta_0 = \lambda(S_0) \) and \( \beta_1 = \lambda(S_1) \), respectively.

**Proposition 7.7.** The walk function \( \gamma_n \) of order \( n \in \mathbb{N} \) for \( (S, \leftrightarrow) \) is given as follows for \( x \in S_i \) and \( i \in \{0, 1\} \):

\[
\begin{align*}
\gamma_n(x) &= \beta_i^{n/2} \beta_{1-i}^{n/2}, & \text{n even} \\
\gamma_n(x) &= \beta_i^{(n-1)/2} \beta_{1-i}^{(n+1)/2}, & \text{n odd}
\end{align*}
\]

**Proof.** Note that a walk in \( (S, \leftrightarrow) \) must alternate between the sets \( S_i \) and \( S_{1-i} \).

**Proposition 7.8.** The unique constant rate distribution for \( (S, \leftrightarrow) \) has rate \( \alpha = 1/\sqrt{\beta_0 \beta_1} \). The density \( f \) of this distribution is given by

\[
\begin{align*}
f(x) &= \frac{1}{\beta_0 + \sqrt{\beta_0 \beta_1}}, & x \in S_0 \\
f(x) &= \frac{1}{\beta_1 + \sqrt{\beta_0 \beta_1}}, & x \in S_1
\end{align*}
\]

**Proof.** The equation for the constant rate in Theorem 7.6 is \( \alpha \beta_0/\alpha + \alpha \beta_1/\alpha + \alpha \beta_1/\alpha + \alpha \beta_0 = 1 \), which simplifies to \( \alpha^2 \beta_0 \beta_1 = 1 \). Hence \( \alpha = 1/\sqrt{\beta_0 \beta_1} \). The form of \( f \) follows from the general results.

**Theorem 7.7.** The random walk \( X = (X_1, X_2, \ldots) \) on \( (S, \leftrightarrow) \) associated with the constant rate distribution has transition density \( P \) given by \( P(x, y) = 1/\beta_1 \) if \( x \in S_0 \), \( y \in S_1 \) and \( P(x, y) = 1/\beta_0 \) if \( x \in S_1 \), \( y \in S_0 \).

**Proof.** This follows easily from the definition \( P(x, y) = f(y)/F(x) \) for \( x \leftrightarrow y \) and Theorem 7.8. The result also follows from the general theory.

So for \( n \in \mathbb{N}_+ \), if \( X_n \in S_0 \) then \( X_{n+1} \) is uniformly distributed on \( S_1 \) and if \( X_n \in S_1 \) then \( X_{n+1} \) is uniformly distributed on \( S_0 \).
Example 7.1 (The star). Consider the star on $k + 1$ vertices where $k \in \mathbb{N}_+$. Denote the center by 0 and the endpoints by \{1, 2, \ldots, k\}. This is a complete bipartite graph with $\beta_0 = 1$ and $\beta_1 = k$. Hence the unique constant rate is $\alpha = 1/\sqrt{k}$ and the corresponding density function $f$ is given by

\[
\begin{align*}
    f(0) &= \frac{1}{1 + \sqrt{k}} \\
    f(x) &= \frac{1}{k + \sqrt{k}}, \quad x \in \{1, 2, \ldots, k\}
\end{align*}
\]
Chapter 8

Product Spaces

In this chapter, we consider various products of graphs and semigroups. Throughout we have $\sigma$-finite measure spaces $(S, \mathcal{F}, \mu)$ and $(T, \mathcal{I}, \nu)$, each with a measurable diagonal, as discussed in Chapter 1. The corresponding product measure space $(S \times T, \mathcal{F} \times \mathcal{I}, \mu \times \nu)$ is also $\sigma$-finite with a measurable diagonal. Of course in the discrete case, the sets are countable, the $\sigma$-algebras are the power sets, and measures are counting measure.

8.1 Direct Product of Graphs

Suppose that $(S, \rightarrow)$ and $(T, \uparrow)$ are graphs.

**Definition 8.1.** The direct product of $(S, \rightarrow)$ and $(T, \uparrow)$ is the graph $(S \times T, \nearrow)$ where
\[(x, y) \nearrow (z, w) \text{ if and only if } x \rightarrow z \text{ and } y \uparrow w\]

**Proof.** We need to note that the appropriate measurability condition holds. Note that the adjacency kernel $L$ of $(S \times T, \nearrow)$ is related to the adjacency kernels $J$ and $K$ of $(S, \rightarrow)$ and $(T, \uparrow)$ by
\[L[(x, y), (z, w)] = J(x, z)K(y, w), \quad (x, y), (z, w) \in S \times T\]
so $L$ is measurable.

In the discrete case, the direct product, as we have defined it, is sometimes referred to as the categorical product or the tensor product. The left walk functions are related as follows:

**Theorem 8.1.** Let $\delta_n, \gamma_n$ and $\rho_n$ denote the left walk function of order $n \in \mathbb{N}_+$ for the graphs $(S, \rightarrow), (T, \uparrow)$ and $(S \times T, \nearrow)$, respectively. Then $\rho_n(x, y) = \delta_n(x)\gamma_n(y)$ for $(x, y) \in S \times T$.

**Proof.** This follows since $((x_1, y_1), (x_2, y_2), \ldots, (x_{n+1}, y_{n+1}))$ is a walk of length $n$ in $(S \times T, \nearrow)$ if and only if $(x_1, x_2, \ldots, x_{n+1})$ and $(y_1, y_2, \ldots, y_{n+1})$ are walks of length $n$ in $(S, \rightarrow)$ and $(T, \uparrow)$ respectively.

Note however that the left generating function of $(S \times T, \nearrow)$ has no simple representation in terms of the left generating functions of $(S, \rightarrow)$ and $(T, \uparrow)$. The left generating functions are
\[
[x, y], t \mapsto \sum_{n=0}^{\infty} \rho_n(x, y)t^n = \sum_{n=0}^{\infty} \delta_n(x)\gamma_n(y)t^n \\
(x, t) \mapsto \sum_{n=0}^{\infty} \delta_n(x)t^n, \quad (x, t) \mapsto \sum_{n=0}^{\infty} \gamma_n(y)t^n
\]

The direct product is the most common way of forming a partial order on a product space from partial orders on the component spaces.

**Theorem 8.2.** If $(S, \leq_1)$ and $(T, \leq_2)$ are partial order graphs then the direct product $(S \times T, \leq)$ is also a partial order graph.
Proof. If \((x, y) \in S \times T\) then \(x \preceq_1 x\) and \(y \preceq_2 y\) so \((x, y) \preceq (x, y)\). Hence \(\preceq\) is reflexive. Suppose \((u, v), (x, y) \in S \times T\) and that \((u, v) \preceq (x, y)\) and \((x, y) \preceq (u, v)\). Then \(u \preceq_1 x\), \(x \preceq_1 u\), \(v \preceq_2 y\), and \(y \preceq_2 v\). Hence \(u = x\) and \(v = y\) so \((u, v) = (x, y)\). Thus \(\preceq\) is antisymmetric. Finally, suppose that \((u, v), (x, y), (z, w) \in S \times T\) and that \((u, v) \preceq (x, y)\) and \((x, y) \preceq (z, w)\). Then \(u \preceq_1 x\), \(x \preceq_1 z\), \(v \preceq_2 y\), and \(y \preceq_2 w\). Hence \(u \preceq_1 z\) and \(v \preceq_2 w\) so \((u, v) \preceq (z, w)\). Thus \(\preceq\) is transitive.

Returning to the general case, suppose that \(X\) and \(Y\) are random variables with values in \(S\) and \(T\), respectively. Let \(F, G,\) and \(H\) denote the right probability functions of \(X, Y\), and \((X, Y)\) for the graphs \((S, \rightarrow)\), \((T, \uparrow)\), and \((S \times T, \nearrow)\), respectively. The following simple result refers to the problem of recovering \(F\) and \(G\) from \(H\).

**Proposition 8.1.** Distribution functions.

(a) If \(e \in S\) satisfies \(e \rightarrow x\) for all \(x \in S\) then \(G(y) = H(e, y)\) for \(y \in T\).

(b) If \(e \in T\) satisfies \(e \uparrow y\) for all \(y \in T\) then \(F(x) = H(x, e)\) for \(x \in S\).

**Proof.** The proofs are trivial.

(a) Note that \(G(y) = P(y \uparrow Y) = P(e \rightarrow X, y \uparrow Y) = H(e, y)\) for \(y \in T\).

(b) Similarly, \(H(x) = P(x \rightarrow X) = P(x \rightarrow X, e \uparrow Y) = H(x, e)\) for \(x \in S\).

Note that \(e\) in part (a) and \(e\) in part (b) are not necessarily unique unless the relations are anti-symmetric. But in particular, (a) holds for a partial order graph \((S, \preceq_1)\) with minimum element \(e \in S\) and (b) holds for a partial order graph \((T, \preceq_2)\) with minimum element \(e\).

**Definition 8.2.** Random variables \(X\) and \(Y\) are right independent for the graphs \((S, \rightarrow)\) and \((T, \uparrow)\) if the events \(\{x \rightarrow X\}\) and \(\{y \uparrow Y\}\) are independent for every \(x \in S\) and \(y \in T\).

Of course, if \(X\) and \(Y\) are independent, then they are right independent. The converse is not generally true. The following result is trivial.

**Proposition 8.2.** Random variables \(X\) and \(Y\) are right independent for the graphs \((S, \rightarrow)\) and \((T, \uparrow)\) respectively if and only if

\[ H(x, y) = F(x)G(y), \quad (x, y) \in S \times T \]

**Proof.** By definition, \(F(x) = P(x \rightarrow X)\) for \(x \in S\), \(G(y) = P(y \uparrow Y)\) for \(y \in T\), and,

\[ H(x, y) = P[(x, y) \nearrow (X, Y)] = P(x \rightarrow X, y \uparrow Y), \quad (x, y) \in S \times T \]

so the result follows immediately.

**Corollary 8.1.** Suppose that \(X\) and \(Y\) are independent random variables with values in \(S\) and \(T\). Suppose also that \(X\) has right rate function \(p\) for \((S, \rightarrow)\) and that \(Y\) has right rate function \(q\) for \((T, \uparrow)\). Then \((X, Y)\) has right rate function \(r\) for \((S \times T, \nearrow)\) where

\[ r(x, y) = p(x)q(y), \quad (x, y) \in S \times T \]

In particular, if \(X\) has constant rate \(\alpha \in (0, \infty)\) for \((S, \rightarrow)\) and \(Y\) has constant rate \(\beta \in (0, \infty)\) for \((T, \uparrow)\) then \((X, Y)\) has constant rate \(\alpha \beta\) for \((S \times T, \nearrow)\).

**Proof.** By assumption, \(X\) and \(Y\) have density functions \(f\) and \(g\) and then \(p = f/F\) and \(q = g/G\) where, as before, \(F\) and \(G\) are the right probability functions of \(X, Y\) for the graphs \((S, \rightarrow)\), \((T, \uparrow)\), respectively. By independence, a density function \(h\) for \((X, Y)\) is given by \(h(x, y) = f(x)g(x)\) for \((x, y) \in S \times T\), and by Theorem 8.2, \((X, Y)\) has right probability function \(H\) for \((S \times T, \nearrow)\) where \(H(x, y) = F(x)G(y)\) for \((x, y) \in S \times T\). Hence \((X, Y)\) has right rate function \(r\) given by

\[ r(x, y) = \frac{h(x, y)}{H(x, y)} = \frac{f(x)g(y)}{F(x)G(y)} = p(x)q(y), \quad x \in S, y \in T \]
Here is a converse.

**Corollary 8.2.** Suppose again that $X$ and $Y$ are random variables with values in $S$ and $T$. If $(X,Y)$ has constant rate $\delta \in (0,\infty)$ for the graph $(S \times T, \nearrow)$ and if $X$ and $Y$ are right independent for $(S, \rightarrow)$ and $(T, \uparrow)$, then $X$ and $Y$ are independent, and $X$ has constant rate $\alpha$ for $(S, \rightarrow)$ and $Y$ has constant rate $\beta$ for $(T, \uparrow)$ for some $\alpha, \beta \in (0,\infty)$ with $\alpha\beta = \delta$.

**Proof.** Let $F$, $G$, and $H$ denote the right probability functions of $X$, $Y$, and $(X,Y)$ for the graphs $(S, \rightarrow)$, $(T, \uparrow)$ and $(S \times T, \nearrow)$, respectively. Then from the constant rate property and Theorem 8.2, $(X,Y)$ has density function $h$ given by

$$h(x, y) = \delta H(x, y) = \delta F(x)G(y), \quad x \in S, \ y \in T$$

By the factorization theorem, it follows that $X$ and $Y$ are independent, and there are density functions $f$ and $g$ of $X$ and $Y$ and positive constants $\alpha$ and $\beta$ such that $f(x) = \alpha F(x)$, $g(y) = \beta G(y)$ for $x \in S$ and $y \in T$ such that $\alpha\beta = \delta$. \hfill \Box

For the remainder of this section, suppose that $(S, \rightarrow)$ is a general graph, as above, but that $(T, \equiv)$ is the complete graph where $v \equiv y$ for every $v, y \in T$. Recall that $(T, \equiv)$ is the graph associated with the right trivial semigroup $(T, \cdot)$. The direct product $(S \times T, \rightarrow)$ of $(S, \rightarrow)$ and $(T, \equiv)$ is useful for counterexamples, and many of the results above simplify considerably. We will assume that $c = \nu(T) \in (0,\infty)$ so that $(T, \equiv)$ is finite.

**Proposition 8.3.** Walk functions and generating functions.

(a) For $n \in \mathbb{N}$, let $\delta_n$ and $\gamma_n$ denote the left walk functions of order $n$ for the graphs $(S, \rightarrow)$ and $(S \times T, \rightarrow)$, respectively. Then

$$\gamma_n(x, y) = e^n \delta_n(x), \quad (x, y) \in S \times T$$

(b) Let $\Delta$ and $\Gamma$ denote the left generating functions for the graphs $(S, \rightarrow)$ and $(S \times T, \Rightarrow)$ respectively. Then for $(x, y) \in S \times T$ and $ct$ in the interval of convergence of $\Delta(x, \cdot)$,

$$\Gamma[(x, y), t] = \Delta(x, ct)$$

**Proposition 8.4.** Suppose that $(X,Y)$ is a random variable with values in $S \times T$.

(a) Let $F$ and $G$ denote the right probability functions for $(S \times T, \Rightarrow)$ and $(S, \rightarrow)$, respectively. Then $F(x, y) = G(x)$ for $(x, y) \in S \times T$.

(b) $X$ and $Y$ are right independent.

(c) $(X,Y)$ has constant rate $\alpha \in (0,\infty)$ for $(S \times T, \Rightarrow)$ if and only if $X$ and $Y$ are independent, $X$ has constant rate $\alpha c$ for $(S, \rightarrow)$, and $Y$ is uniformly distributed on $T$.

We will return to this setting later in this chapter in Section 8.3 on the direct product of semigroups.

### 8.2 Cartesian Products of Graphs

Suppose that $(S, \rightarrow)$ and $(T, \uparrow)$ are discrete, irreflexive graphs, and so are graphs without loops in the ordinary combinatorial sense.

**Definition 8.3.** The **Cartesian product** of $(S, \rightarrow)$ and $(T, \uparrow)$ is the graph $(S \times T, \nearrow)$ where $\nearrow$ is defined by

$$(x, y) \nearrow (z, w) \text{ if and only if } (x \rightarrow z \text{ and } y = w) \text{ or } (x = z \text{ and } y \uparrow w)$$

The Cartesian product is a well known construction in graph theory. The theorem and corollary below are our main results.

**Theorem 8.3.** Suppose that $X$ and $Y$ are independent random variables with values in $S$ and $T$ and with probability density functions $f$ and $g$, respectively. Let $F$ and $G$ denote the right probability functions of $X$ and $Y$ with respect to the graphs $(S, \rightarrow)$ and $(S_2, \uparrow)$, respectively. Then $(X,Y)$ has right probability function $H$ with respect to $(S, \nearrow)$ given by

$$H(x, y) = F(x)g(y) + f(x)G(y), \quad (x, y) \in S$$
Proof. By definition of the relation and by independence,
\[ H(x, y) = P[(x, y) \rightarrow (X, Y)] = P(x \rightarrow X, Y = y) + P(X = x, y \uparrow Y) \]
\[ = P(x \rightarrow X)P(Y = y) + P(X = x)P(y \uparrow Y) = F(x)g(y) + f(x)G(y) \] \[(x, y) \in S \]
We have used the fact that the events \( \{x \rightarrow X, Y = y\} \) and \( \{X = x, y \uparrow Y\} \) are disjoint, since the graphs are irreflexive.

**Corollary 8.3.** Suppose that again that \( X \) and \( Y \) are independent random variables with values in \( S \) and \( T \), and that \( X \) has right rate function \( p \) for \( (S, \rightarrow) \) and that \( Y \) has right rate function \( q \) for \( (T, \uparrow) \). Then \((X, Y)\) has right rate function \( r \) for \((S \times T, \rightarrow)\) given by

\[ r(x, y) = \frac{p(x)q(y)}{p(x) + q(y)} \]

In particular, if \( X \) has constant rate \( \alpha \) for \((S, \rightarrow)\) and \( Y \) has constant rate \( \beta \) for \((T, \uparrow)\) then \((X, Y)\) has constant rate \( \alpha \beta / (\alpha + \beta) \) for \((S \times T, \rightarrow)\).

**Proof.** As before, let \( f, g, \) and \( h \) denote the density functions of \( X, Y, \) and \((X, Y)\), and let \( F, G, \) and \( H \) denote the right probability functions of \( X, Y, \) and \((X, Y)\) in terms of their respective graphs. Then

\[ r(x, y) = \frac{h(x, y)}{H(x, y)} = \frac{f(x)g(y)}{F(x)g(y) + f(x)G(y)} \]

Dividing the numerator and denominator by \( F(x)G(y) \) gives the result.

**Note 8.1.** In the context and language of spectral graph theory, parts of the theorem and corollary are well known.

**Note 8.2.** Suppose that \((S, \rightarrow)\) and \((T, \uparrow)\) are discrete, irreflexive graphs. Suppose again that \( X \) and \( Y \) are independent random variables with values in \( S \) and \( T \) respectively, and that \( X \) has constant rate \( \alpha \) for \((S, \rightarrow)\) and \( Y \) has constant rate \( \beta \) for \((S, \uparrow)\). Combining the results in this section and the last we have the interesting result that \((X, Y)\) has constant rate \( \alpha \beta / (\alpha + \beta) \) for the direct product \((S \times T, \rightarrow)\) and has constant rate \( \alpha \beta / (\alpha + \beta) \) for the Cartesian product \((S \times T, \rightarrow)\).

### 8.3 Direct Products of Semigroups

Next we consider semigroups relative to our underlying measure spaces \((S, \mathcal{S}, \mu)\) and \((T, \mathcal{S}, \nu)\). We will use \( \cdot \) (and concatenation) generically as the semigroup operator, regardless of the underlying set, but the semigroup under consideration should be clear from context.

**Definition 8.4.** Suppose that \((S, \cdot)\) and \((T, \cdot)\) are semigroups relative to the two measure spaces. The **direct product** is the semigroup \((S \times T, \cdot)\) with the binary operation \( \cdot \) defined by

\[ (x, y)(u, v) = (xu, yv) \]

**Proof.** We need to show that the product space is a semigroup and satisfies the assumptions we have imposed. Let \((u, v), (x, y), (z, w) \in S \times T\). Then

\[ [(x, y)(u, v)](z, w) = (xu, yv)(z, w) = (xuz, yvw) = (x, y)(uw, zv) = (x, y)(u, v)(z, w) \]

so the associative property holds. Next suppose that \((u, v)(x, y) = (u, v)(z, w)\). Then \(ux = uz\) and \(vy = vw\). Hence \(x = z\) and \(y = w\) so \((x, y) = (z, w)\). Therefore the left cancellation law holds. Finally,

\[ [(x, y), (u, v)] \mapsto (x, y)(u, v) = (xu, yv) \]

is measurable since \((x, u) \mapsto xu\) and \((y, v) \mapsto yv\) are measurable. \(\square\)

Note that if \((x, y) \in S \times T, A \in \mathcal{S}, \) and \(B \in \mathcal{S}\) then

\[ (x, y)(A \times B) = (xA) \times (yB) \]

\[ (x, y)^{-1}(A \times B) = (x^{-1}A) \times (y^{-1}B) \]
**Theorem 8.4.** Suppose again that \((S, \cdot)\) and \((T, \cdot)\) are semigroups with associated graphs \((S, \to)\) and \((T, \uparrow)\), respectively. Then the graph \((S \times T, \rhd)\) associated with the product semigroup \((S \times T, \cdot)\) is the direct product of the graphs \((S, \to)\) and \((T, \uparrow)\).

**Proof.** Let \((u, v), (x, y) \in S \times T\). By definition, \((u, v) \rhd (x, y)\) if and only if \((x, y) = (s, t)(u, v)\) for some \((s, t) \in S \times T\). In turn, this is true if and only if \(x = su\) and \(y = tv\) so \(u \to x\) and \(v \uparrow y\).

So all of the results in Section 8.1 on direct products of graphs apply to direct products of semigroups.

**Theorem 8.5.** Suppose again that \((S, \cdot)\) and \((T, \cdot)\) are semigroups with direct product \((S \times T, \cdot)\).

(a) If \((S, \cdot)\) and \((T, \cdot)\) are positive semigroups, then so is \((S \times T, \cdot)\).

(b) If \((S, \cdot)\) or \((T, \cdot)\) is a strict positive semigroup, then so is \((S \times T, \cdot)\).

**Proof.** The proofs are straightforward.

(a) Let \(e \in S\) and \(e \in T\) denote the identity elements of \((S, \cdot)\) and \((T, \cdot)\) respectively. Then for \((x, y) \in S \times T\), \((x, y)(e, e) = (xe, ye) = (x, y)\) and \((e, e)(x, y) = (ex, ey) = (x, y)\). So \((e, e)\) is the identity for \((S \times T, \cdot)\).

(b) Suppose again that \((x, y), (u, v) \in S \times T\). Then \((u, v)(x, y) = (ux, vy)\). Since one of the semigroups is strictly positive, either \(ux \neq u\) or \(vy \neq v\). In both cases, \((u, v)(x, y) \neq (u, v)\).

In part (b), the strict positive semigroup \((S \times T, \cdot)\) can be made into a positive semigroup with the addition of an identity element \((e, e)\) as described in Theorem 3.7. For the next theorem, recall the definition of the right trivial semigroup in Example 3.1

**Theorem 8.6.** If \((S, \cdot)\) and \((T, \cdot)\) are right trivial semigroups then so is the direct product \((S \times T, \cdot)\).

**Proof.** By definition, \((u, v)(x, y) = (ux, vy) = (x, y), (u, v), (x, y) \in S \times T\)

**Theorem 8.7.** If \(\mu\) and \(\nu\) are left-invariant measures for \((S, \cdot)\) and \((T, \cdot)\), respectively, then \(\mu \times \nu\) is left invariant for \((S \times T, \cdot)\).

**Proof.** For \(x \in S\), \(y \in T\) and \(A \in \mathcal{F}\), and \(B \in \mathcal{G}\),

\[(\mu \times \nu)(\{x, y\} \times \{A, B\}) = (\mu \times \nu)(x A \times y B)\]

\[= \mu(x A) \nu(y B) = \mu(A) \nu(B) = (\mu \times \nu)(A \times B)\]

Therefore, for fixed \((x, y) \in S \times T\), the measures on \(S \times T\)

\[C \mapsto (\mu \times \nu)(\{x, y\} \times C)\]

\[C \mapsto (\mu \times \nu)(C)\]

agree on the measurable rectangles \(A \times B\) where \(A \in \mathcal{F}\) and \(B \in \mathcal{G}\). Hence, these measures must agree on all of \(\mathcal{F} \times \mathcal{G}\), and hence \(\mu \times \nu\) is left-invariant for \((S \times T, \cdot)\).

Suppose now that \((S, \cdot)\) and \((T, \cdot)\) are positive semigroups with identity elements \(e\) and \(e\), respectively, and that the left-invariant measures \(\mu\) and \(\nu\) are unique, up to multiplication by positive constants. We show that \(\mu \times \nu\) has the same property. Let \(\mathcal{F}(T) = \{B \in \mathcal{F} : \nu(T) \in (0, \infty)\}\). Suppose now that \(\lambda\) is a left-invariant measure for \((S \times T, \cdot)\). For \(C \in \mathcal{F}(T)\), define

\[\mu_C(A) = \lambda(A \times C), \quad A \in \mathcal{F}\]

Then \(\mu_C\) is a regular measure on \(S\) (although it may not have support \(S\)). Moreover, for \(x \in S\) and \(A \in \mathcal{F}\),

\[\mu_C(x A) = \lambda(x A \times C) = \lambda((x, e)(A \times C)) = \lambda(A \times C) = \mu_C(A)\]
so $\mu_C$ is left-invariant for $(S, \cdot)$. It follows that for each $C \in \mathcal{B}(T)$, there exists $\rho(C) \in [0, \infty)$ such that $\mu_C = \rho(C)\mu$; that is,

$$\lambda(A \times C) = \mu(A)\rho(C), \quad A \in \mathcal{B}(S), \ C \in \mathcal{B}(T) \quad (8.1)$$

Fix $A \in \mathcal{S}$ with $\mu(A) \in (0, \infty)$. If $C, D \in \mathcal{B}(T)$ and $C \subseteq D$ then

$$\mu(A)\rho(C) = \lambda(A \times C) \leq \lambda(A \times D) = \mu(A)\rho(D)$$

so $\rho(C) \leq \rho(D)$. If $C, D \in \mathcal{B}(T)$ are disjoint then

$$\mu(A)(C \cup D) = \lambda(A \times (C \cup D)) = \lambda((A \times C) \cup (A \times D))$$

$$= \lambda(A \times C) + \lambda(A \times D) = \mu(A)\rho(C) + \mu(A)\rho(D)$$

so $\rho(C \cup D) = \rho(C) + \rho(D)$. If $C, D \in \mathcal{B}(T)$ then

$$\mu(A)\rho(C \cup D) = \lambda(A \times (C \cup D)) = \lambda((A \times C) \cup (A \times D))$$

$$\leq \lambda(A \times C) + \lambda(A \times D) = \mu(A)\rho(C) + \mu(A)\rho(D)$$

so $\rho(C \cup D) \leq \rho(C) + \rho(D)$. Thus, $\rho$ is a content in the sense of $[19]$, and hence can be extended to a regular measure on $T$ (which we will continue to call $\rho$). Thus from (8.1) we have

$$\lambda(A \times C) = (\mu \times \rho)(A \times C), \quad A \in \mathcal{S}, \ B \in \mathcal{B}(T)$$

By regularity, it follows that $\lambda = \mu \times \rho$. Again fix $A \in \mathcal{S}$ with $0 < \mu(A) < \infty$. If $y \in T$ and $B \in \mathcal{B}$ then

$$\mu(A)\rho(yB) = \lambda(A \times yB) = \lambda((e, y)(A \times B)) = \lambda(A \times B) = \mu(A)\rho(B)$$

so it follows that $\rho(yB) = \rho(B)$ and hence $\rho$ is left-invariant for $(T, \cdot)$. Thus, $\rho = c\nu$ for some positive constant $c$ and so $\lambda = c(\mu \times \nu)$. Therefore $\mu \times \nu$ is the unique left-invariant measure for $(S \times T, \cdot)$, up to multiplication by positive constants.

For the two theorems and the corollary that follow, we assume that $(S, \cdot)$ and $(T, \cdot)$ are positive semigrons with identity elements $e \in S$ and $\epsilon \in T$, respectively, so that $(S \times T, \cdot)$ is also a positive semigroup. Suppose that $X$ and $Y$ are random variables with values in $S$ and $T$ respectively. Our interest is in the memoryless and exponential properties.

**Theorem 8.8.** $(X, Y)$ is memoryless for $(S \times T, \cdot)$ if and only if $X$ is memoryless for $(S, \cdot)$, $Y$ is memoryless for $(T, \cdot)$, and $X, Y$ are right independent.

**Proof.** Let $F$, $G$, and $H$ denote the right probability functions for $X$, $Y$, and $(X, Y)$ with respect to the semigroups $(S, \cdot)$, $(T, \cdot)$ and $(S \times T, \cdot)$ respectively. Suppose first that $(X, Y)$ is memoryless for $(S \times T, \cdot)$. The from Theorem 8.1,

$$F(ux) = H(ux, e) = H[(u, e)(x, e)] = H(u, e)H(xe) = F(u)F(x), \quad u, x \in S$$

So $X$ is memoryless for $(S, \cdot)$ By a symmetric argument, $Y$ is memoryless for $(T, \cdot)$. Next note that

$$H(x, y) = H[(x, e)(e, y)] = H(x, e)H(e, y) = F(x)G(y), \quad x \in S, \ y \in T$$

so $X$ and $Y$ are right independent. Conversely, suppose that $X$ and $Y$ are memoryless and are right independent. Then

$$H[(u, v)(x, y)] = H(ux, vy) = F(ux)G(vy) = F(u)F(x)G(v)G(y)$$

$$= [F(u)G(v)][F(x)G(y)] = H(u, v)H(x, y), \quad (u, v), (x, y) \in S \times T$$

Hence $(X, Y)$ is memoryless for $(S \times T, \cdot)$.

**Theorem 8.9.** If $(X, Y)$ is exponential for $(S \times T, \cdot)$ then $X$ is exponential for $(S, \cdot)$, $Y$ is exponential for $(T, \cdot)$, and $X$ and $Y$ are right independent. Conversely, if $X$ and $Y$ are exponential for $(S, \cdot)$ and $(T, \cdot)$ respectively and are independent, then $(X, Y)$ is exponential for $(S \times T, \cdot)$.
Proof. Suppose that \((X, Y)\) is exponential for \((S \times T, \cdot)\). Then
\[
\mathbb{P}(X \in xA) = \mathbb{P}[(X, Y) \in xA \times T] = \mathbb{P}[(X, Y) \in (x, \epsilon)(A \times T)]
\]
\[
= \mathbb{P}[(X, Y) \in (x, \epsilon)(S \times T)]\mathbb{P}[(X, Y) \in A \times T] = \mathbb{P}(X \in xS)\mathbb{P}(X \in A), \quad x \in S, \ A \in \mathcal{S}
\]
Hence \(X\) is exponential for \((S, \cdot)\). By a symmetric argument, \(Y\) is exponential for \((T, \cdot)\). Finally, since \((X, Y)\) is exponential for \((S \times T, \cdot)\), it is also memoryless and hence \(X\) and \(Y\) are right independent by Theorem 8.8. Conversely, suppose that \(X\) and \(Y\) are independent and are exponential for \((S, \cdot)\) and \((T, \cdot)\) respectively. If \(A \in \mathcal{S}, \ B \in \mathcal{S}, \) and \((x, y) \in S \times T\) then
\[
\mathbb{P}[(X, Y) \in (x, y)(A \times B)] = \mathbb{P}[X \in xA, Y \in yB] = \mathbb{P}(X \in xA)\mathbb{P}(Y \in yB)
\]
\[
= \mathbb{P}(X \in xS)\mathbb{P}(X \in A)\mathbb{P}(Y \in yT)\mathbb{P}(Y \in B) = \mathbb{P}(X \in xS, Y \in yT)\mathbb{P}(X \in A, Y \in B)
\]
\[
= \mathbb{P}[(X, Y) \in (x, y)(S \times T)]\mathbb{P}[(X, Y) \in A \times B], \quad (x, y) \in S \times T
\]
Hence for fixed \((x, y) \in S \times T\), the finite measures on \(\mathcal{S} \times \mathcal{S}\) given by
\[
C \mapsto \mathbb{P}[(X, Y) \in (x, y)C]
\]
\[
C \mapsto \mathbb{P}[(X, Y) \in (x, y)S]\mathbb{P}[(X, Y) \in C]
\]
agree on the measurable rectangles \(A \times B\) where \(A \in \mathcal{S}\) and \(B \in \mathcal{S}\). Hence these measures agree on \(\mathcal{S} \times \mathcal{S}\) and so \((X, Y)\) is exponential for \((S \times T, \cdot)\).

In the first half of the proof, note that \((X, Y)\) has constant rate \(\delta \in (0, \infty)\) for \((S \times T, \cdot)\) with respect to a left invariant measure \(\lambda\) on \((S \times T, \mathcal{S} \times \mathcal{S})\). But we cannot use Corollary 8.2 to conclude that \(X\) and \(Y\) are independent since we don’t know that \(\lambda\) is a product measure on \((S \times T, \mathcal{S} \times \mathcal{S})\). Note that the canonical such measure \(\lambda\) is given by
\[
\lambda(A \times B) = \mathbb{E} \left[ \frac{1}{H(X, Y)}; (X, Y) \in A \times B \right] = \mathbb{E} \left[ \frac{1}{F(X)G(Y)}; X \in A, Y \in B \right]
\]
But we cannot factor the expression further without full independence of \(X\) and \(Y\). However, we have the following corollary:

Corollary 8.4. Suppose that \(\lambda = \mu \times \nu\) is the unique left-invariant measure for \((S \times T, \mathcal{S} \times \mathcal{S})\), up to multiplication by positive constants, as in Theorem 8.7. Then \((X, Y)\) is exponential for \((S \times T, \cdot)\) if and only if \(X\) is exponential for \((S, \cdot)\), \(Y\) is exponential for \((T, \cdot)\), and \(X\) and \(Y\) are independent.

The direct product \((S \times T, \cdot)\) has several natural sub-semigroups. First, \(\{(x, \epsilon) : x \in S\}\) is a complete sub-semigroup isomorphic to \(S\). Similarly, \(\{(\epsilon, y) : y \in T\}\) is a complete sub-semigroup isomorphic to \(T\). If \(S = T\), then the diagonal \(\{(x, x) : x \in S\}\) is a complete sub-semigroup isomorphic to \(S\).

Naturally, the results in this section can be extended to the direct product of \(n \in \mathbb{N}_+\) positive semigroups \((S_1, \cdot), (S_2, \cdot), \ldots, (S_n, \cdot)\) and in particular to the \(n\)-fold direct power \((S^n, \cdot)\) of a positive semigroup \((S, \cdot)\). In the latter case, if \(\lambda\) is left invariant for \((S, \cdot)\) then \(\lambda^n\) is left invariant for \((S^n, \cdot)\) for each \(n \in \mathbb{N}_+\). This fact was used in the proof of Proposition 3.1. For an infinite construction, suppose that \((S_i, \cdot)\) is a discrete positive semigroup with identity element \(e_i\) for \(i \in \mathbb{N}_+\). We can construct an infinite product that will be quite useful. Let
\[
T = \{(x_1, x_2, \ldots) : x_i \in S_i \text{ for each } i \text{ and } x_i = e_i \text{ for all but finitely many } i\}
\]
As before, we define the component-wise operation:
\[
(x_1, x_2, \ldots) \cdot (y_1, y_2, \ldots) = (x_1y_1, x_2y_2, \ldots)
\]
Then \((T, \cdot)\) is a discrete positive semigroup.

Our final result is useful for counterexamples. We assume that \((S, \cdot)\) is a general semigroup and that \((T, \cdot)\) is a right trivial semigroup. Recall that the graph associated with \((T, \cdot)\) is the complete graph \((T, \equiv)\) so that \(y \equiv v\) for every \((y, v) \in T^2\). In particular, Proposition 8.3 and Proposition 8.4 in Section 8.1 on the direct product of graphs hold. Note also that the previous two theorems and the corollary in this section do not apply, since \((S, \cdot)\) may not be a positive semigroup, and \((T, \cdot)\) certainly is not. Nonetheless, similar results hold.
**Proposition 8.5.** Suppose that \((X,Y)\) is a random variable with values in \(S \times T\).

(a) \((X,Y)\) is memoryless for \((S \times T, \cdot)\) if and only if \(X\) is memoryless for \((S, \cdot)\).

(b) If \((X,Y)\) is exponential for \((S \times T, \cdot)\) then \(X\) is exponential for \((S, \cdot)\).

(c) Conversely, if \(X\) and \(Y\) are independent and \(X\) is exponential for \((S, \cdot)\) then \((X,Y)\) is exponential for \((S \times T, \cdot)\).

*Proof.* The main point to remember is that \(yB = B\) for \(y \in T\) and \(B \in \mathcal{F}\). and so in particular, if \(G\) denotes the right probability function of \(X\) for \((S, \cdot)\) then the right probability function \(F\) for \((S \times T, \cdot)\) is given by \(F(x,y) = G(x)\) for \((x,y) \in S \times T\).

(a) Let \((u,v), (x,y) \in S \times T\). Then \(F[(u,v)(x,y)] = F(uv, xy) = G(u)\). Also \(F(u,v) = G(u)\) and \(F(x,y) = G(x)\).

(b) Let \(x \in S, y \in T, \text{ and } A \in \mathcal{F}\). Then

\[
\mathbb{P}(X \in xA) = \mathbb{P}(X \in xA, Y \in T) = \mathbb{P}(X \in xA, Y \in yT) = \mathbb{P}[(X,Y) \in (x,y)(A \times T)]
\]

\[
= F(x,y)\mathbb{P}[(X,Y) \in A \times T] = G(x)\mathbb{P}(X \in A)
\]

(c) Let \((x,y) \in S \times T\) and let \(A \in \mathcal{F}, B \in \mathcal{F}\). Then

\[
\mathbb{P}[(X,Y) \in (x,y)(A \times B)] = \mathbb{P}(X \in xA, Y \in yB) = \mathbb{P}(X \in xA)\mathbb{P}(Y \in yB)
\]

\[
= G(x)\mathbb{P}(X \in A)\mathbb{P}(Y \in B) = F(x,y)\mathbb{P}(X \in A, Y \in B)
\]

\[
= F(x,y)\mathbb{P}[(X,Y) \in A \times B]
\]

It then follows more generally that \(\mathbb{P}[(X,Y) \in (x,y)C] = F(x,y)\mathbb{P}[(X,Y) \in C]\) for \(C \in \mathcal{F} \times \mathcal{F}\).

\[\square\]

This construction provides examples of semigroups that have distributions that are memoryless or even exponential, but do not have constant rate for the underlying graph. (Of course, \((T, \cdot)\) has this property also, but the product semigroup is a little less trivial.) Specifically, suppose that \(\mu\) and \(\nu\) are fixed reference measures for \((S, \mathcal{F})\) and \((T, \mathcal{F})\) and that \(c = \nu(T) < \infty\) so that the graph \((T, \equiv)\) is finite. Suppose now that \(X\) and \(Y\) are independent random variables with values in \(S\) and \(T\), respectively. Suppose also that \(X\) is exponential for \((S, \cdot)\) but \(Y\) is not uniformly distributed on \(T\). Then \((X,Y)\) is exponential for \((S \times T, \cdot)\) but does not have constant rate for the corresponding graph \((S \times T, \rightarrow)\) (the direct product of \((S, \rightarrow)\) and \((T, \equiv))\).

### 8.4 Lexicographic Sums of Graphs

For the final section of this chapter, we modify the underlying measure spaces. Suppose that \((S, \rightarrow)\) is a discrete, irreflexive graph (so a graph without loops in the combinatorial sense) and that \((T_x, \uparrow_x)\) is a general graph for each \(x \in S\). Underlying \((S, \rightarrow)\) is the discrete measure space of course, and underlying \((T_x, \uparrow_x)\) is a \(\sigma\)-finite measure space \((T_x, \mathcal{F}_x, \mu_x)\), as usual. Let \(T = \bigcup_{x \in S} T_x\). Here is the new space we will need:

**Definition 8.5.** Define the \(\sigma\)-finite measure space \((S \bullet T, \mathcal{F} \bullet \mathcal{F}, \lambda)\) as follows:

(a) \(S \bullet T = \bigcup_{x \in S} \{x\} \times T_x\)

(b) \(\mathcal{F} \bullet \mathcal{F} = \{\bigcup_{x \in S} \{x\} \times A_x : A_x \in \mathcal{F}_x\text{ for all } x \in S\}\).

(c) \(\lambda[\bigcup_{x \in S} \{x\} \times A_x] = \sum_{x \in S} \mu_x(A_x)\) where \(A_x \in \mathcal{F}_x\) for \(x \in S\).

Note that \(S \bullet T \subseteq S \times T\). In the special case that \((T_x, \mathcal{F}_x, \lambda_x)\) is the common measure space \((T, \mathcal{F}, \mu)\) for all \(x \in S\), the measure space \((S \bullet T, \mathcal{F} \bullet \mathcal{F}, \lambda)\) is the product space \((S \times T, \mathcal{F} \times \mathcal{F}, \# \times \mu)\) where \(\mathcal{F}\) is the power set of \(S\) and where \(#\) is counting measure as usual. Next is the graph we will study.
Definition 8.6. The lexicographic sum of the graphs \((T_x, \uparrow_x)\) over \((S, \rightarrow)\) is the graph \((S \bullet T, \searrow)\) where
\[(u, v) \searrow (x, y) \text{ if and only if } u \rightarrow x \text{ or } (u = x \text{ and } v \uparrow_x y)\]
In the special case that \((T_x, \uparrow_x)\) is the common graph \((T, \uparrow)\) (on the common measure space) for all \(x \in S\), the graph \((S \times T, \searrow)\) is the lexicographic product of \((S, \rightarrow)\) and \((T, \uparrow)\).

The lexicographic sum (and particularly the lexicographic product) is a common construction in ordinary combinatorial graph theory. It is also common in the study of partial orders, and it is in this context that the graph gets its name.

Theorem 8.10. Suppose that \((S, \prec_0)\) is a discrete, strict partial order graph.

(a) If \((T_x, \prec_x)\) is a strict partial order graph for each \(x \in S\), then the lexicographic sum \((S \bullet T, \prec)\) is also a strict partial order graph:
\[(u, v) \prec (x, y) \text{ if and only if } u \prec_0 x \text{ or } (u = x \text{ and } v \prec_x y)\]

(b) If \((T, \preceq_x)\) is a partial order graph for each \(x \in S\), then the lexicographic sum \((S \bullet T, \preceq)\) is also a partial order graph:
\[(u, v) \preceq (x, y) \text{ if and only if } u \prec_0 x \text{ or } (u = x \text{ and } v \preceq_x y)\]

Proof. The proofs are simple from the various cases in the definition.

(a) First \((x, y) \not\prec (x, y)\) for \(x, y \in T_x\), so \(\prec\) is irreflexive. If \((u, v) \prec (x, y)\) and \((x, y) \prec (z, w)\) then \((u, v) \prec (z, w)\) for \((u, v), (x, y), (z, w) \in S \bullet T\), so \(\prec\) is transitive.

(b) It’s clear that \(\preceq\) is the partial order associated with the strict partial order \(\prec\) in (a), when \(\preceq_x\) is the partial order associated with the strict partial order \(\prec_x\) for each \(x \in S\).

\[\square\]

Returning to the general setting, there is a simple relationship between the left walk functions of order 1. The relationship is more complicated for higher orders, but the order 1 relationship is important for the constant rate distributions that we will discuss below. To setup the notation, let \(\delta, \sigma_x\) for \(x \in S\), and \(\rho\) denote the left walk functions of order 1 for \((S, \rightarrow)\), \((T_x, \uparrow_x)\) for \(x \in S\), and \((U, \searrow)\), respectively.

Theorem 8.11. The left walk functions of order 1 are related as follows:
\[\rho(x, y) = \sum_{u \rightarrow x} \mu_u(T_u) + \sigma_x(y), \quad x \in S, y \in T_x\]

Proof. This follows from the definition of the left walk function.
\[\rho(x, y) = \lambda\{(u, v) \in U : (u, v) \searrow (x, y)\}\]
\[= \lambda\{(u, v) : u \in S, v \in T_u, u \rightarrow x\} + \lambda\{(x, v) : v \in T_x, v \uparrow y\}\]
\[= \sum_{u \rightarrow x} \mu_u(T_u) + \sigma_x(y), \quad x \in S, y \in T_x\]

Note that \(\rho(x, y) < \infty\) implies \(\sigma_x(y) < \infty\) and \(\sum_{u \rightarrow x} \mu_u(T_u) < \infty\). Conversely, if \(\sigma(y) < \infty, \delta(x) < \infty\) and \(\mu_u(T_u) < \infty\) for every \(u \rightarrow x\) then \(\rho(x, y) < \infty\). In the special case that \((S \times T, \searrow)\) is the lexicographic product of \((S, \rightarrow)\) and \((T, \uparrow)\), the result in Theorem 8.11 becomes
\[\rho(x, y) = \mu(T)\delta(x) + \sigma(y), \quad x \in S, y \in T\]

Suppose now that \((X, Y)\) is a random variable with values in \(S \bullet T\), so that \(X\) has values in \(S\), and given \(X = x\), random variable \(Y\) has values in \(T_x\). Unconditionally, \(Y\) has values in \(T = \bigcup_{x \in S} T_x\). To setup the notation, let \(F\) denote the right probability function of \(X\) for \((S, \rightarrow)\), and for \(x \in S\), let \(G(\cdot \mid x)\) denote the conditional right probability function of \(Y\) for \((T_x, \uparrow_x)\), given that \(X = x\). Let \(H\) denote the right probability function of \((X, Y)\) for \((U, \searrow)\). Finally, let \(f\) denote the density function of \(X\), so that \(f(x) = \mathbb{P}(X = x)\) for \(x \in S\). In addition to our usual support assumptions, we assume that \(f(x) > 0\) for \(x \in S\).
**Theorem 8.12.** The right probability functions are related as follows:

\[ H(x, y) = F(x) + f(x)G(y \mid x), \quad x \in S, \; y \in T_x \]

**Proof.** This follows from the definition of the lexicographic relation \(\triangleleft\):

\[
H(x, y) = \mathbb{P}[(x, y) \, \triangleleft (X, Y)] = \mathbb{P}(X = x, y \uparrow Y) = \mathbb{P}(x \rightarrow X) + \mathbb{P}(X = x)\mathbb{P}(y \uparrow Y \mid X = x) = F(x) + f(x)G(y \mid x), \quad x \in S, \; y \in T_x
\]

Note that since the discrete graph \((S, \rightarrow)\) is irreflexive, the events \(\{x \rightarrow X\}\) and \(\{X = x\}\) are disjoint.

The reason for requiring that the graph \((S, \rightarrow)\) be discrete is clear. If \(X\) has a continuous distribution on \(S\) then \(\mathbb{P}(X = x) = 0\) for \(x \in S\) and so the lexicographic relation is irrelevant from a probabilistic viewpoint.

**Corollary 8.5.** Consider the special case of the lexicographic product \((S \times T, \triangleleft)\) of \((S, \rightarrow)\) and \((T, \uparrow)\). In this case, if \(X\) and \(Y\) are independent, then we have

\[ H(x, y) = F(x) + f(x)G(y), \quad x \in S, \; y \in T \]

where \(G\) is the right probability function of \(Y\) for \((T, \uparrow)\).

Suppose that given \(X = x \in S\), random variable \(Y\) has density function \(g(\cdot \mid x)\) on \(T_x\). Then \((X, Y)\) has density \(h\) given by

\[ h(x, y) = f(x)g(y \mid x) \quad \text{for } x \in S, \; y \in T_x. \]

Hence the rate function of \((X, Y)\) for \((U, \triangleleft)\) is given by

\[ (x, y) \mapsto \frac{f(x)g(y \mid x)}{F(x) + f(x)G(y \mid x)}, \quad x \in S, \; y \in T_x \tag{8.2} \]

As usual, we have special interest in constant rate distributions. Recall that the existence of a constant rate distribution for a graph implies that the left walk function is finite.

**Theorem 8.13.** Suppose that \(X\) has constant rate \(\alpha \in (0, \infty)\) for \((S, \rightarrow)\), and that given \(X = x \in S\), the conditional density function of \(Y\) and the conditional right probability function of \(Y\) for \((T_x, \uparrow_x)\) are related by

\[ g(y \mid x) = \beta[1 + \alpha G(y \mid x)], \quad y \in T_x \]

for some \(\beta \in (0, \infty)\). Then \((X, Y)\) has constant rate \(\alpha \beta\) for \((T, \triangleleft)\).

**Proof.** The assumptions are \(f(x) = \alpha F(x)\) for \(x \in S\) and \(g(y \mid x) = \beta[1 + \alpha G(y \mid x)]\) for \(x \in S\) and \(y \in T_x\). The result then follows by substitution in (8.2).

The following example shows that the conditions are not vacuous.

**Example 8.1.** Consider the special case of the lexicographic sum \((S \bullet T, \triangleleft)\) of the discrete graphs \((T_x, =)\) over the discrete, irreflexive graph \((S, \rightarrow)\). So \((u, v) \, \triangleleft (x, y)\) if and only if either \(u \rightarrow x\) or \((u, v) = (x, y)\). The left walk function relation in Theorem 8.11 becomes

\[ \rho(x, y) = \sum_{u \rightarrow x} \#(T_u) + 1, \quad x \in T, \; y \in T \]

If \((X, Y)\) is a random variable with values in \(S \bullet T\), then given \(X = x \in S\), the conditional density function of \(Y\) and the conditional right probability function of \(Y\) for \((T_x, =)\) are the same:

\[ G(y \mid x) = g(y \mid x) = \mathbb{P}(Y = y \mid X = x), \quad y \in T_x \]

Suppose now that \(\#(T_x) = k \in \mathbb{N}_+\) for all \(x \in S\) so that the sets all have the same cardinality. Suppose also that \(X\) has constant rate \(\alpha \in (0, \infty)\) and that given \(X = x \in S\), random variable \(Y\) is uniformly distributed over \(T_x\). Then the condition in Theorem 8.13 is satisfied with \(\beta = 1/(\alpha + k)\). Hence \((X, Y)\) has constant rate \(\alpha/(\alpha + k)\) for \((U, \triangleleft)\).

**Theorem 8.14.** Suppose that \((X, Y)\) has constant rate \(\alpha \in (0, \infty)\) for \((S \bullet T, \triangleleft)\). Then

(a) The density function \(f\) of \(X\) is given by

\[ f(x) = \frac{\alpha F(x)\mu_x(S_x)}{1 - \alpha \mathbb{E}[\sigma_x(Y) \mid X = x]}, \quad x \in S \]
(b) For \(x \in S\), a conditional density of \(Y\) given \(X = x\) is defined by
\[
g(y \mid x) = \alpha \left[ \frac{F(x)}{f(x)} + G(y \mid x) \right], \quad y \in T_x
\]

(c) \(Y\) has density function \(g\) defined by
\[
g(y) = \alpha \left( \mathbb{E}[\delta(X)] + \mathbb{E}[G(y \mid X)] \right), \quad y \in T
\]

**Proof.** By assumption, \(h = \alpha H\) is a probability density function of \((X, Y)\).

(a) This follows the basic moment result in Corollary 4.2:
\[
f(x) = \int_{T_x} h(x, y) \, d\mu_x(y) = \alpha F(x) \mu_x(T_x) + \alpha f(x) \int_{T_x} G(y \mid x) \, d\mu_x(y)
\]

\[
= \alpha F(x) \mu_x(T_x) + \alpha f(x) \mathbb{E}[\sigma_x(Y) \mid X = x], \quad x \in S
\]

Solving for \(f(x)\) gives the result.

(b) For \(x \in S\),
\[
g(y \mid x) = \frac{h(x, y)}{f(x)} = \alpha \frac{F(x) + f(x) G(y \mid x)}{f(x)}, \quad y \in T_x
\]

Simplifying gives the result.

(c) Again we use the basic moment result in Corollary 4.2.
\[
g(y) = \sum_{x \in S} h(x, y) = \alpha \sum_{x \in S} F(x) + \alpha \sum_{x \in S} f(x) G(y \mid x) = \alpha \mathbb{E}[\delta(X)] + \alpha \mathbb{E}[G(y \mid X)], \quad y \in T
\]

\[\square\]

**Corollary 8.6.** Suppose that \((S \times T, \preceq)\) is the lexicographic product of \((S, \rightarrow)\) and \((T, \uparrow)\) and that \((X, Y)\) has constant rate \(\alpha\) for \((S \times T, \preceq)\).

(a) If \(X\) and \(Y\) are independent, then \(X\) has constant rate \(\beta\) for \((S, \rightarrow)\) where
\[
\beta = \frac{\alpha \mu(T)}{1 - \alpha \mathbb{E}[\sigma(Y)]}
\]

(b) \(Y\) does not have constant rate for \((T, \uparrow)\) except in the special case that \((T, \uparrow)\) is right regular and \(Y\) is uniformly distributed on \(T\).

**Example 8.2.** The standard graph \(([0, \infty), \leq)\) is isomorphic to the lexicographic product \((\mathbb{N} \times [0, 1), \leq)\) of \((\mathbb{N}, <)\) and \(([0, 1), \leq)\). (Of course, \([0, \infty)\) and \([0, 1)\) have the usual Lebesgue measure structure.) An isomorphism from the product graph to the standard graph is \((n, t) \mapsto n + t\) where \(n \in \mathbb{N}\) and \(t \in [0, 1)\). The inverse, an isomorphism from the standard graph to the product graph, is \(x \mapsto ([x], x - [x])\) for \(x \in [0, \infty)\).

In the context of Corollary 8.6, note that \(\mu(S) = 1\), \(\delta(n) = n - 1\) for \(n \in \mathbb{N}_+\) (and \(\delta(0) = 0\), and \(\sigma(y) = y\) for \(y \in [0, 1)\).

Suppose now that \(X\) has the exponential distribution on \([0, \infty)\) with parameter \(\alpha \in (0, \infty)\). Of course, \(X\) has constant rate \(\alpha\) for the graph \(([0, \infty), \leq)\). Since the spaces are isomorphic, \((N, Y)\) must have constant rate \(\alpha\) for the lexicographic product graph, where \(N = \lfloor X \rfloor\) is the integer part of \(X\) and \(Y = X - N\) is the remainder. As is well known (or easy to see directly), \(N\) has the geometric distribution on \(\mathbb{N}\) with density function \(f\) given by
\[
f(n) = (1 - e^{-\alpha}) e^{-\alpha n}, \quad n \in \mathbb{N}
\]

The right probability function \(F\) of \(N\) for the graph \((\mathbb{N}, <)\) is given by
\[
F(n) = e^{-\alpha(n+1)}, \quad n \in \mathbb{N}
\]
So $N$ has constant rate $\beta = (1 - e^{-\alpha})/e^{-\alpha}$ for $(\mathbb{N}, \prec)$. On the other hand, the distribution of $Y$ is the same as the conditional distribution of $X$ given $X \in [0, 1)$, with density function $g$ given by

$$g(y) = \frac{\alpha}{1 - e^{-\alpha}} e^{-\alpha y}, \quad y \in [0, 1)$$

The right probability function $G$ of $Y$ for the graph $([0, \infty), \leq)$ is given by

$$G(y) = \int_y^1 \frac{\alpha}{1 - e^{-\alpha}} e^{-\alpha t} dt = \frac{1}{1 - e^{-\alpha}} e^{-\alpha y} - \frac{e^{-\alpha}}{1 - e^{-\alpha}}, \quad y \in [0, 1)$$

It’s also well known (or easy to see directly) that $N$ and $Y$ are independent. So the density function $h$ of $(N, Y)$ is given by

$$h(n, y) = f(n) g(y) = \alpha e^{-\alpha n} e^{-\alpha y} = \alpha e^{-\alpha (n+y)}, \quad (n, y) \in \mathbb{N} \times [0, 1)$$

The right probability function $H$ of $(N, Y)$ for the product graph $(\mathbb{N} \times [0, 1), \preceq)$ is given by

$$H(n, y) = F(n) + f(n) G(y) = e^{\alpha(n+1)} + (1 - e^{-\alpha}) e^{-\alpha n} \frac{1}{1 - e^{-\alpha}} (e^{-\alpha y} - e^{-\alpha})$$

$$= e^{-\alpha n} e^{-\alpha y} = e^{-\alpha (n+y)}, \quad (n, y) \in \mathbb{N} \times [0, 1)$$

So $(N, Y)$ has constant rate $\alpha$ for $(\mathbb{N} \times [0, 1), \preceq)$ as of course it must. In the context of Theorem 8.13, simple computations show that

$$\mathbb{E}[\delta(N)] = \frac{e^{-\alpha}}{1 - e^{-\alpha}}, \quad \mathbb{E}[\sigma(Y)] = \frac{1}{\alpha} - \frac{e^{-\alpha}}{1 - e^{-\alpha}}$$

We can then verify that

$$\beta = \frac{\alpha \mu(T)}{1 - \alpha \mathbb{E}[\sigma(Y)]}$$

and that

$$g(y) = \alpha \mathbb{E}[\delta(N)] + G(y), \quad y \in [0, 1)$$

consistent with Corollary 8.14. We will return to this basic example, with additional structure, in the next chapter.
Chapter 9

Quotient spaces

9.1 Preliminaries

As usual, we start with a measurable space \((S, \mathcal{S})\) satisfying the diagonal property. Suppose that \((S, \cdot)\) is a positive semigroup with identity element \(e\) and that \((T, \cdot)\) is a measurable, positive sub-semigroup of \((S, \cdot)\).

Let \(\preceq\) denote the relation on \(S\) associated with \(S\) itself, and \(\preceq_T\) the relation on \(S\) associated with \(T\). So for \((x, y) \in S^2\), \(x \preceq y\) if and only if \(y \in xS\), and \(x \preceq_T y\) if and only if \(y \in xT\). Recall that \(\preceq\) and \(\preceq_T\) are partial orders on \(S\), and \(\preceq_T\) is a sub order of \(\preceq\). That is, \(x \preceq_T y\) implies \(x \preceq y\) for \((x, y) \in S^2\). In this section, we will use the notation

\[ [e, x]_T = \{ t \in T : t \preceq x \}, \quad x \in S \]

Recall also that \(S_+ = S - \{e\}\) and similarly \(T_+ = T - \{e\}\).

**Definition 9.1.** Define the quotient space of \(S\) by \(T\) to be the set \(S/T = \{ z \in S : [e, z]_T = \{e\} \}\).

That is, \(z \in S/T\) if and only if \(t \preceq z\) imply \(t = e\).

**Proposition 9.1.** \(T \cap (S/T) = \{e\}\).

**Proof.** First \(e \in T \cap (S/T)\) since \(e \in T\), and trivially if \(t \in T\) and \(t \preceq e\) then \(t = e\). Conversely, suppose that \(t \in T \cap (S/T)\). Since \(t \preceq e\), it follows that \(t = e\). \(\square\)

**Proposition 9.2.** Suppose that \(x \in S\). Then \(x = yz\) for some \(y \in T\) and \(z \in S/T\) if and only if \(y\) is a maximal element of \([e, x]_T\) with respect to \(\preceq_T\) (and then \(z = y^{-1}x\)).

**Proof.** Suppose that \(x = yz\) for some \(y \in T\) and \(z \in S/T\). Then \(y \preceq x\) by definition, so \(y \in [e, x]_T\). Suppose that \(t \in [e, x]_T\) and \(y \preceq_T t\). There exists \(a \in S\) and \(b \in T\) such that \(x = ta\) and \(t = yb\). Hence \(x = yba\). By the left cancellation rule, \(z = ba\), so \(b \preceq z\). But \(z \in S/T\), so \(b = e\) and hence \(t = y\). Therefore \(y\) is a maximal element of \([e, x]_T\) with respect to \(\preceq_T\). Conversely, suppose that \(y\) is a maximal element of \([e, x]_T\) with respect to \(\preceq_T\). Then \(y \preceq x\) so \(x = yz\) for some \(z \in S\). Suppose that \(t \in T\) and \(t \preceq z\). Then \(z = tb\) for some \(b \in S\) so \(x = ytb\). Hence \(yt \preceq x\) and \(yt \in T\). Since \(y\) is maximal, \(yt = y\) and so \(t = e\). Therefore \(z \in S/T\). \(\square\)

**Proposition 9.3.** \(S/T = \bigcap_{t \in T_+} (S - tS)\)

**Proof.** Suppose that \(z \in S/T\) and let \(t \in T_+\). If \(z \in tS\), then \(t \preceq z\) which implies \(t = e\), a contradiction. Hence \(z \in S - tS\). Conversely, suppose that \(z \in S - tS\) for every \(t \in T_+\). Then \(t \not\preceq z\) for every \(t \in T_+\) and hence \(t \in T\) and \(t \preceq z\) imply \(t = e\). Therefore \(z \in S/T\). \(\square\)

For the remainder of this section, we impose the following assumptions:

**Assumption 9.1.** For each \(x \in S\), \([e, x]_T\) has a unique maximal element \(\varphi_T(x)\) (with respect to \(\preceq_T\)). The function \(\varphi_T : S \to T\) is measurable.

Thus \(S/T = \{ z \in S : \varphi_T(z) = e \}\), so \(S/T\) is measurable as well. For \(x \in S\), let \(\psi_T(x) = \varphi_T^{-1}(x)x \in S/T\) so that \(x \in S\) can be factored uniquely as \(x = \varphi_T(x)\psi_T(x)\). This quotient space structure does correspond to an equivalence relation: if we define \(u \sim v\) if and only if \(\psi_T(u) = \psi_T(v)\), then \(\sim\) is an equivalence relation on \(S\) and the elements in \(S/T = \text{range}(\psi_T)\) generate a complete set of equivalence classes. However, \(\sim\) is
not a congruence in the sense of [17]. That is, if \( u \sim v \), it is not necessarily true that \( xu \sim xv \) (unless, of course, \( x \in T \)). Finally, note that the mapping \((\varphi_T, \psi_T) : S \rightarrow T \times (S/T)\) is one-to-one and onto. The measure-theoretic structure is preserved as well, under a basic assumption. As usual, we give \( T \) and \( S/T \) the induced \( \sigma \)-algebras \( \mathcal{F} = \{ A \in \mathcal{F} : A \subseteq T \} \) and \( \mathcal{F}/\mathcal{T} = \{ A \in \mathcal{F} : A \subseteq S/T \} \).

**Assumption 9.2.** Assume that \( \lambda \) is a \( \sigma \)-finite measure on \((S, \mathcal{F})\) that is left invariant for \((S, \cdot)\), and that \( \mu \) is a \( \sigma \)-finite measure on \((T, \mathcal{F})\) that is left invariant for \((T, \cdot)\), both unique up to multiplication by positive constants.

**Theorem 9.1.** There exists a measure \( \nu \) on \((S/T, \mathcal{F}/\mathcal{T})\) such that

\[
\lambda(AB) = \mu(A)\nu(B), \quad A \in \mathcal{F}, \ B \in \mathcal{F}/\mathcal{T}
\]

*Proof.* Let \( \mathcal{C} = \{ C \in \mathcal{F}/\mathcal{T} : \lambda(C) \in (0, \infty) \} \). For \( C \in \mathcal{C} \) and \( A \in \mathcal{F} \), let \( \mu_C(A) = \lambda(AC) \). Then \( \mu_C \) is a regular measure on \((T, \mathcal{F})\) for each \( C \in \mathcal{C} \) (although \( \mu_C \) may not have support \( T \)). Moreover, for \( y \in T \) and \( A \in \mathcal{F} \),

\[
\mu_C(yA) = \lambda(yAC) = \lambda(AC) = \mu_C(A)
\]

so \( \mu_C \) is left-invariant for \((T, \cdot)\). It follows that for each \( C \in \mathcal{C} \), there exists \( \nu(C) \in [0, \infty) \) such that \( \mu_C = \nu(C)\mu \); that is,

\[
\lambda(AC) = \mu(A)\nu(C), \quad A \in \mathcal{F}, \ C \in \mathcal{C}
\]

Fix \( A \in \mathcal{F} \) with \( \mu(A) \in (0, \infty) \). If \( C, \ D \in \mathcal{F}/\mathcal{T} \) with \( C \subseteq D \) then

\[
\mu(A)\nu(C) = \lambda(AC) \leq \lambda(AD) = \mu(A)\nu(D)
\]

and hence \( \nu(C) \leq \nu(D) \). If \( C, \ D \in \mathcal{C} \) are disjoint then \( AC \) and \( AD \) are also disjoint and hence

\[
\mu(A)\nu(C \cup D) = \lambda[AC(D)] = \lambda(AC \cup AD)
\]

\[
= \lambda(AC) + \lambda(AD) = \mu(A)\nu(C) + \mu(A)\nu(D)
\]

Therefore \( \nu(C \cup D) = \nu(C) + \nu(D) \). Finally, if \( C, \ D \in \mathcal{C} \) then

\[
\mu(A)\nu(C \cup D) = \lambda[AC(D)] = \lambda(AC \cup AD)
\]

\[
\leq \lambda(AC) + \lambda(AD) = \mu(A)\nu(C) + \mu(A)\nu(D)
\]

and so \( \nu(C \cup D) \leq \nu(C) + \nu(D) \). It follows that \( \nu \) is a content on \((S/T, \mathcal{F}/\mathcal{T})\) in the sense of [19], and hence can be extended to a measure on \((S/T, \mathcal{F}/\mathcal{T})\) (which we will also call \( \nu \)). It now follows from regularity that

\[
\lambda(AB) = \mu(A)\nu(B), \quad A \in \mathcal{F}, \ B \in \mathcal{F}/\mathcal{T}
\]

\[\square\]

The following examples should help clarify the assumptions.

**Example 9.1.** Let \((S, \cdot)\) be a positive semigroup. For \( t \in S_+ \) let \( S_t = \{ t^n : n \in \mathbb{N} \} \). Then \((S_t, \cdot)\) is a complete, positive sub-semigroup of \((S, \cdot)\) and is clearly isomorphic to the standard, discrete positive semigroup \((\mathbb{N}, +)\), under the mapping \( n \mapsto t^n \). The quotient space is

\[
S/S_t = \bigcap_{n=1}^{\infty} (S - t^nS) = S - tS = \{ z \in S : t \nmid z \}
\]

Suppose that \( \{ n \in \mathbb{N} : t^n \preceq x \} \) is finite for each \( x \in S \). The set is clearly nonempty since \( t^0 = e \preceq x \). Hence the set has a maximum element

\[
n_t(x) = \max\{ n \in \mathbb{N} : t^n \preceq x \}
\]

So the basic assumptions are satisfied and \( \varphi_t(x) = t^{n_t(x)} \) for \( x \in S \). That is, each \( x \in S \) has a unique factoring as

\[
x = t^nz, \quad n \in \mathbb{N}, \ z \in S - tS
\]

Suppose that \( \lambda \) is a left-invariant measure for \((S, \cdot)\). Of course, counting measure \# is left invariant for \((S_t, \cdot)\). The decomposition in Theorem 9.1 holds with \( \nu \) being the restriction of \( \lambda \) to \( S/S_t \). That is

\[
\lambda(AB) = #(A)\lambda(B), \quad A \subseteq S_t, \ B \subseteq S - tS
\]
Example 9.2. Consider the standard positive semigroup \((0,\infty)^k, +\). The associated order \(\leq\) is the ordinary (product) order. Let \(T = \mathbb{N}^k\), so that \((T,+)\) is a discrete, complete, positive sub-semigroup of \((S,+)\). The quotient space is \(S/T = [0,1)^k\) and the assumptions are satisfied. Each \(x \in [0,\infty)^k\) can be decomposed uniquely as
\[
x = n + t, \quad n \in \mathbb{N}^k, \quad t \in [0,1)^k
\]
where \(n\) is the vector of integer parts of \(x\) and \(t\) is the vector of remainders. The left-invariant measure on \([0,\infty)^k\) is \(k\)-dimensional Lebesgue measure \(\lambda\), and the left-invariant measure on \(T\) is counting measure \#.

The reference measure on \(S/T\) is also \(k\)-dimensional Lebesgue measure. Moreover, the partial order graph \((S,\leq)\) is the lexicographic product of \((T,\prec)\) with \((S/T,\leq)\), as discussed in Section 8.4.

Example 9.3. Consider the direct product \((S,\cdot)\) of positive semigroups \((S_1,\cdot)\) and \((S_2,\cdot)\), with identity elements \(e_1\) and \(e_2\), and with left-invariant measures \(\lambda_1\) and \(\lambda_2\), respectively, as discussed in Section 8.3. Let
\[
T_1 = \{(x_1,e_2) : x_1 \in S_1\}
\]
Then \(T_1\) is a complete positive sub-semigroup of \(S\) satisfying the assumptions. Moreover, the quotient space
\[
T_2 := S/T_1 = \{(e_1,x_2) : x_2 \in S_2\}
\]
is also a complete positive sub-semigroup. In this example, the spaces \(T_1\) and \(T_2 = S/T_1\) are symmetric; \(T_1\) is isomorphic to \(S_1\) and \(T_2\) is isomorphic to \(S_2\). The unique factoring is simply \((x_1,x_2) = (x_1,e_2)(e_1,x_2)\). The measures \(\mu\) and \(\nu\) in Theorem 9.1 are given by
\[
\mu(A) = \lambda_1(A_1), \quad A \in \mathcal{F}_1
\]
\[
\nu(B) = \lambda_2(B_2), \quad B \in \mathcal{F}_2
\]
where \(A_1 = \{x_1 \in S_1 : (x_1,e_2) \in A\}\) and \(B_2 = \{x_2 \in S_2 : (e_1,x_2) \in B\}\).

9.2 Distributions

We return to the general setting of a positive semigroup \((S,\cdot)\) on an underlying measurable space \((S,\mathcal{S})\). As usual, let \(\preceq\) denote the partial order associated with \((S,\cdot)\). Suppose next that \(T\) is a sub-semigroup of \(S\) and that \(X\) is a random variable taking values in \(S\) with \(\mathbb{P}(X \in T) > 0\). Recall from Theorem 5.12 that if \(X\) has an exponential distribution on \((S,\cdot)\) then the conditional distribution of \(X\) given \(X \in T\) is exponential on \((T,\cdot)\), and the right probability function of \(X\) given \(X \in T\) is the restriction to \(T\) of the right probability function of \(X\):
\[
\mathbb{P}(X \succeq_T x \mid X \in T) = \mathbb{P}(X \succeq x), \quad x \in T
\]
We will generalize and extend this basic result. Suppose that \(T\) has a quotient space \(S/T\) as in Definition 9.1, and with Assumptions 9.1 and 9.2 in place. Then random variable \(X\) can be decomposed uniquely as \(X = Y_T Z_T\) where \(Y_T\) takes values in \(T\), and \(Z_T\) takes values in \(S/T\). Our goal is the study of the random variables \(Y_T\) and \(Z_T\). When \(T = S_i\) for \(t \in S - \{\varepsilon\}\), as in Example 9.1, we simplify the notation to \(Y_t\) and \(Z_t\).

In this case, note that \(Y_t = t^{N_t}\) where \(N_t\) takes values in \(\mathbb{N}\). The following theorem is our first main result.

Theorem 9.2. Suppose that \(X\) has an exponential distribution on \((S,\cdot)\). Then
(a) \(Y_T\) has an exponential distribution on \((T,\cdot)\).
(b) The right probability function of \(Y_T\) for \((T,\cdot)\) is the restriction to \(T\) of the right probability function of \(X\) for \((S,\cdot)\).
(c) \(Y_T\) and \(Z_T\) are independent.

Proof. Let \(y \in T\), \(A \in \mathcal{F}\), and \(B \in \mathcal{F}/\mathcal{F}\). Then by the uniqueness of the factorization and since \(X\) has an exponential distribution,
\[
\mathbb{P}(Y_T \in y A, Z_T \in B) = \mathbb{P}(X \in y A B) = \mathbb{P}(X \succeq y) \mathbb{P}(X \in A B)
= \mathbb{P}(X \succeq y) \mathbb{P}(Y_T \in A, Z_T \in B)
\]  (9.1)
$Y_T$.

Chapter 9. Quotient Spaces

86

Substituting $A = T$ and $B = S/T$ gives

$$P(Y_T \geq T \ y) = P(X \geq y), \ y \in T$$

so the right probability function of $Y$ is the restriction to $T$ of the right probability function of $X$. Returning to (9.1) with general $A$ and $B = S/T$ we have

$$P(Y_T \in y \ A) = P(Y_T \geq T \ y)P(Y_T \in A), \ y \in T, \ A \in \mathcal{F}$$

so $Y_T$ has an exponential distribution on $T$. Finally, returning to (9.1) with $A = T$ and general $B$ we have

$$P(Y_T \geq y, Z_T \in B) = P(Y_T \geq T \ y)P(Z_T \in B), \ y \in T, B \in \mathcal{F}/\mathcal{F}$$

so $Y_T$ is right independent of $Z_T$. To get full independence, recall that $X$ has constant rate $\alpha$ with respect to $\lambda$, so $X$ has density $f$ with respect to $\lambda$ given by $f(x) = \alpha P(X \geq x)$ for $x \in S$. Similarly, $Y_T$ has constant rate with respect to $\mu$ on $T$. Thus the function $g$ on $S \times S/T$ given by $g(y, z) = f(yz)$ is a density function of $(Y_T, Z_T)$ with respect to $\mu \times \nu$. By the memoryless property,

$$g(y, z) = f(yz) = \alpha P(X \geq yz) = \alpha P(X \geq y)P(X \geq z), \ y \in S, z \in S/T$$

and so by the standard factorization theorem, $Y_T$ and $Z_T$ are independent. □

Example 9.4. Consider the direct product of positive semigroups $(S_1, \cdot)$ and $(S_2, \cdot)$ with the sub-semigroup and quotient space described in Example 9.3. In this case, the theorem gives another proof of the characterization of exponential distributions in Corollary 8.4: $(X_1, X_2)$ is exponential on $S_1 \times S_2$ if and only if $X_1$ is exponential on $S_1$, $X_2$ is exponential on $S_2$, and $X_1, X_2$ are independent.

Theorem 9.3. Suppose again that $X$ is a random variable taking values in $S$, with the factoring $X = Y_T Z_T$ where $Y_T \in T$ and $Z_T \in S/T$.

(a) If $P(X \in T) > 0$ then the conditional distribution of $X$ given $X \in T$ is the same as the distribution of $Y_T$ if and only if $Y_T$ and $\{Z_T = \epsilon\}$ are independent.

(b) If $P(X \in S/T) > 0$ then the conditional distribution of $X$ given $X \in S/T$ is the same as the distribution of $Z_T$ if and only if $Z_T$ and $\{Y_T = \epsilon\}$ are independent.

Proof. Suppose that $A \in \mathcal{F}$. Then $\{X \in A\} = \{Y_T \in A, Z_T = \epsilon\}$ and in particular, $\{X \in T\} = \{Z_T = \epsilon\}$. Thus,

$$P(X \in A \ | \ X \in T) = P(Y_T \in A \ | \ Z_T = \epsilon)$$

The proof of the second result is analogous. □

Corollary 9.1. Suppose that $X$ has an exponential distribution on $(S, \cdot)$.

(a) If $P(X \in T) > 0$ then the conditional distribution of $X$ given $X \in T$ is the same as the distribution of $Y_T$, and this distribution is exponential on $T$.

(b) If $P(X \in S/T) > 0$ then the conditional distribution of $X$ given $X \in S/T$ is the same as the distribution of $Z_T$.

For the following corollary, recall the sub-semigroup $S_x = \{x^n : n \in \mathbb{N}\}$ for $x \in S_+$.

Example 9.1. We assume that $\{n \in \mathbb{N} : x^n \preceq y\}$ is finite for each $x \in S_+$ and $y \in S$, and hence has a maximum element $n_x(y)$. In this case, the quotient space is $S/S_x = S - xS$ and the factoring for a random variable $X$ is $X = x^{n_x} Z_x$ where $n_x \in \mathbb{N}$ and $Z_x \in S - xS$.

Corollary 9.2. Suppose that $X$ has an exponential distribution on $(S, \cdot)$. Then

(a) $N_x$ has the geometric distribution on $N$ with success parameter $p_x = 1 - P(X \geq t)$ for $x \in S_+$.

(b) $(1 - p_x)(1 - p_y) = 1 - p_{xy}$ for $x, y \in S_+$.

(c) $N_x$ and $Z_x$ are independent for $x \in S_+$.

Proof. Most of the proof follows directly from Theorem 9.2.
(a) Let \( x \in S_+ \). Then \( Y_x \) has an exponential distribution on \((S_x, \cdot)\) and therefore \( N_x \) has a geometric distribution on \( \mathbb{N} \), since \((S_x, \cdot)\) and \((\mathbb{N}, +)\) are isomorphic. Next, \( N_x \geq 1 \) if and only if \( Y_x \geq x \) if and only if \( X \geq x \). Thus, the rate (or success) parameter of the geometric distribution is

\[ p_x = 1 - \mathbb{P}(N_x \geq 1) = 1 - \mathbb{P}(X \geq x) \]

so that \( \mathbb{P}(N_x = n) = p_x(1 - p_x)^n \) for \( n \in \mathbb{N}_+ \).

(b) By the memoryless property,

\[ 1 - p_{xy} = \mathbb{P}(X \geq xy) = \mathbb{P}(X \geq x)\mathbb{P}(X \geq y) = (1 - p_x)(1 - p_y), \quad x, y \in S_+ \]

(c) \( Y_x \) and \( Z_x \) are independent, and hence \( N_x \) and \( Z_x \) are independent for \( x \in S_+ \).

Of course, as noted in the proof, the geometric distribution on \( \mathbb{N} \) with success parameter \( p_x \) is the exponential distribution on \((\mathbb{N}, +)\) with rate parameter \( p_x \). The following theorem is our second main result, and gives a partial converse to Theorem 9.2.

**Theorem 9.4.** Suppose \( X \) is a random variable taking values in \( S \), and that \( N_x \) and \( Z_x \) are independent and \( N_x \) has a geometric distribution on \( \mathbb{N} \) for each \( x \in S_+ \). Then \( X \) has an exponential distribution on \((S, \cdot)\).

**Proof.** Let \( p_x \) denote the parameter of the geometric distribution of \( N_x \), for \( x \in S_+ \) so that

\[ \mathbb{P}(N_x = n) = p_x(1 - p_x)^n, \quad n \in \mathbb{N} \]

Let \( A \in \mathcal{F} \). Then

\[ A = \bigcup_{n=0}^{\infty} [A \cap \{ y \in S : n_x(y) = n \}] = \bigcup_{n=0}^{\infty} x^n B_n \]

where \( B_n = \{ z \in S/S_x : x^nz \in A \} = \{ x^{-n}y : n_x(y) = n, y \in A \} \). The collection \( \{ x^n B_n : n \in \mathbb{N} \} \) is disjoint. Similarly, \( xA = \bigcup_{n=0}^{\infty} x^{n+1} B_n \) and the collection \( \{ x^{n+1} B_n : n \in \mathbb{N} \} \) is disjoint. From the hypotheses,

\[ \mathbb{P}(X \in xA) = \sum_{n=0}^{\infty} \mathbb{P}(X \in x^{n+1} B_n) = \sum_{n=0}^{\infty} \mathbb{P}(N_x = n + 1, Z_x \in B_n) \]

\[ = \sum_{n=0}^{\infty} p_x(1 - p_x)^{n+1} \mathbb{P}(Z_x \in B_n). \]

But also \( 1 - p_x = \mathbb{P}(N_x \geq 1) = \mathbb{P}(Y_x \geq t) = \mathbb{P}(X \geq x) \) so

\[ \mathbb{P}(X \geq x)\mathbb{P}(X \in A) = (1 - p_x) \sum_{n=0}^{\infty} \mathbb{P}(X \in x^n B_n) \]

\[ = (1 - p_x) \sum_{n=0}^{\infty} \mathbb{P}(N_x = n, Z_x \in B_n) \]

\[ = (1 - p_x) \sum_{n=0}^{\infty} p_x(1 - p_x)^n \mathbb{P}(Z_x \in B_n). \]

If follows that \( \mathbb{P}(X \in xA) = \mathbb{P}(X \geq x)\mathbb{P}(X \in A) \) and hence \( X \) has an exponential distribution on \( S \). \( \square \)

As a consequence of Corollary 9.2, \( 1 - p_{xy} = (1 - p_x)(1 - p_y) \) for \( x, y \in S_+ \). Here is a related result with different hypotheses and a slightly weaker conclusion.

**Theorem 9.5.** Suppose again that \( X \) is a random variable with values in \( S \), and that \( N_x \) has a geometric distribution on \( \mathbb{N} \) with success parameter \( p_x \) for \( x \in S_+ \). Suppose also that

\[ 1 - p_{xy} = (1 - p_x)(1 - p_y), \quad x, y \in S_+ \]

Then \( X \) has a memoryless distribution for \((S, \cdot)\).

**Proof.** Once again, \( \mathbb{P}(X \geq t) = \mathbb{P}(Y_t \geq t) = \mathbb{P}(N_t \geq 1) = 1 - p_t \) for \( t \in S_+ \). Hence \( X \) is memoryless because of the condition imposed on \( \{ p_x : x \in S_+ \} \). \( \square \)
Part III

Examples and Applications
Chapter 10

Standard Continuous Spaces

10.1 The standard space

The standard space \([0, \infty), +\) with the ordinary Euclidean measure structure is a positive semigroup. The identity element is 0 and the corresponding partial order is the ordinary total order \(\leq\). Our reference measure is Lebesgue measure \(\lambda\), the only invariant measure, up to multiplication by positive constants. Of course, this is the setting of classical reliability theory, with \([0, \infty)\) representing continuous time, and so is one of the main motivations for the general theory presented in this text. For the most part, the results presented in this section are very well known and need no proofs.

The left walk function \(\gamma_n\) of order \(n \in \mathbb{N}\) for \([0, \infty),\leq\) is given by

\[
\gamma_n(x) = \frac{x^n}{n!}, \quad x \in [0, \infty)
\]

Hence the left generating function \(\Lambda\) of \([0, \infty), \leq\) is given by

\[
\Lambda(x, t) = \sum_{n=0}^{\infty} \frac{x^n}{n!} t^n = e^{tx}; \quad x \in [0, \infty), t \in \mathbb{R}
\]

The semigroup \((S, +)\) and the ordered space \((S, \leq)\) have dimension 1.

Suppose that \(X\) is a random variable with values in \([0, \infty)\), thought of as the lifetime of a device or an organism. The right probability function \(F\) of \(X\) for \([0, \infty), \leq\) is the ordinary reliability function:

\[
F(x) = P(X \geq x), \quad x \in [0, \infty)
\]

So \(F(x)\) is the probability that the device lasts at least until time \(x \in [0, \infty)\). If \(X\) has a continuous distribution with density function \(f\) then the right rate function of \(X\) for \([0, \infty), \leq\) is the usual failure rate function \(r = f/F\). So roughly, \(r(x)\,dx\) is the probability that the device will fail in the small time interval from \(x\) to \(x + dx\), given survival up to time \(x\). The basic moment result becomes

\[
\int_0^{\infty} \frac{x^n}{n!} F(x) \, dx = \mathbb{E} \left[ \frac{X^{n+1}}{(n+1)!} \right], \quad n \in \mathbb{N}
\]

The distribution of \(X\) is memoryless for \([0, \infty), +)\) if and only if it is exponential for \((0, \infty), \leq\) if and only if it has constant rate for \([0, +), \leq\). The memoryless property has the familiar form

\[
P(X \geq x + y \mid X \geq x) = P(X \geq y), \quad x, y \in [0, \infty)
\]

so as long as the device has not failed, it is just as good as new. The exponential distribution with constant rate \(\alpha \in (0, \infty)\) has density function \(f\) and right probability function \(F\) given by

\[
f(x) = \alpha e^{-\alpha x}, \quad F(x) = \alpha e^{-\alpha x}, \quad x \in [0, \infty)
\]

Of course, this is the ordinary exponential distribution on \([0, \infty)\) with parameter \(\alpha\). In this case, \(X\) maximizes entropy over all random variables with \(\mathbb{E}(Y) = \mathbb{E}(X) = 1/\alpha\); the maximum entropy is

\[
H(X) = 1 - \ln \alpha
\]
It is also well known that \( X \) has a compound Poisson distribution.

Suppose now that \( X = (X_1, X_2, \ldots) \) is the random walk on \((0, \infty), +\), or equivalently \([0, \infty], \leq\), associated with a distribution on \([0, \infty)\). Then \( X \) can be constructed as the partial sum process corresponding to an IID sequence \( U = (U_1, U_2, \ldots) \) with the common distribution. That is, \( X_n = \sum_{i=1}^{n} U_i \). Equivalently, \( X \) can be constructed as the sequence of record values associated with \( U \). That is, \( X_1 = U_1 \) and then \( X_n \) is the \( n \)th record in the sequence \( U \) for \( n \in \{2, 3, \ldots\} \).

Suppose that \( X \) is the random walk associated with the exponential distribution with rate \( \alpha \in (0, \infty) \). The transition density \( P \) of \( X \) as is given by

\[
P(x, y) = \alpha e^{-(y-x)}, \quad 0 \leq x \leq y
\]

For \( n \in \mathbb{N}_+ \), the sequence \((X_1, X_2, \ldots, X_n)\) has density \( h_n \) defined by

\[
h_n(x_1, x_2, \ldots, x_n) = \alpha^n e^{-x_n}, \quad 0 \leq x_1 \leq x_2 \leq \cdots \leq x_n
\]

and random variable \( X_n \) has probability density function \( f_n \) defined by

\[
f_n(x) = \alpha^n x^{n-1} (\alpha x)^{n-1} (n-1)! e^{-\alpha x}, \quad x \in [0, \infty)
\]

Of course, this is the ordinary gamma distribution with parameters \( n \) and \( \alpha \).

Let \( N = \{N_x : x \in [0, \infty)\} \) denote the counting process associated with the random walk \( X \), so that \( N_x = \#\{n \in \mathbb{N}_+ : X_n \leq x\} \) for \( x \in [0, \infty) \). Then \( N \) is the ordinary Poisson counting process. To check our results, we will compute the renewal function via the formula in Corollary 4.13

\[
m(x) = E(N_x) = E[\Gamma(X, \alpha), X \leq x] = E(e^{\alpha X}, X \leq x)
\]

\[
= \int_0^x t \alpha e^{-\alpha t} dt = \alpha x, \quad x \in [0, \infty)
\]

We also check the thinning result from Theorem 4.14: With thinning parameter \( r \in (0, 1) \), the probability density function \( g \) of \( X_N \), the first accepted point is given by

\[
g(x) = r \alpha \Gamma(x, (1-r)\alpha) F(x) = r \alpha e^{(1-r)\alpha} e^{-\alpha x} = r \alpha e^{-r \alpha x}, \quad x \in [0, \infty)
\]

so \( X_N \) has the exponential distribution on \((0, \infty), +\) with rate \( r \alpha \).

For \( t \in (0, \infty) \), the sub-semigroup generated by \( t \) is \( \{nt : n \in \mathbb{N}\} \) and the corresponding quotient space is \([0, t)\). The basic assumption in Example 9.1 is satisfied since \( \{n \in \mathbb{N} : nt \leq x\} \) is finite for every \( x \in [0, \infty) \).

So \( x \in [0, \infty) \) can be written uniquely as

\[
x = tn_t(x) + z_t(x)
\]

where \( n_t(x) = \lfloor x/t \rfloor \in \mathbb{N} \) and \( z_t(x) = x - tn_t(x) = x \mod t \in [0, t) \). and of course \( x \mapsto n_t(x) \) is measurable.

The following result is a summary of Theorem 9.2 and Corollary 9.1 in this setting.

**Theorem 10.1.** Suppose that \( X \) has the exponential distribution on \((0, \infty), +\) with rate parameter \( \alpha \in (0, \infty) \). For \( t \in (0, \infty) \), we can write \( X = tN_t + Z_t \) where \( N_t \) takes values in \( \mathbb{N} \) and \( Z_t \) takes values in \([0, t)\).

(a) \( N_t \) and \( Z_t \) are independent.

(b) \( N_t \) has the exponential distribution on \((\mathbb{N}, +)\) (that is, the geometric distribution) with rate parameter \( p_t = 1 - e^{-\alpha t} \).

(c) The distribution of \( Z_t \) is the same as the conditional distribution of \( X \) given \( X < t \) and has density function \( s \mapsto \alpha e^{-\alpha s} / (1 - e^{-\alpha t}) \) on \([0, t)\).

In this standard setting, we can do better than the general converse stated earlier. Suppose that \( X \) is a random variable taking values in \([0, \infty) \). Galambos & Kotz [9] (see also [1]) show that if \( N_t \) has a geometric distribution for all \( t \in (0, \infty) \) then \( X \) has an exponential distribution. We now explore a converse based on independence properties of \( N_t \) and \( Z_t \). Suppose that \( X \) has a continuous distribution on \([0, \infty) \) with density function \( f \) and with right probability function \( F \) for \((0, \infty), \leq\). If \( Z_t \) and \( \{N_t = 0\} \) are independent for each \( t \in (0, \infty) \) then we can assume (by an appropriate choice of the density function) that

\[
f(s) = [1 - F(t)] \sum_{n=0}^{\infty} f(nt + s)
\]

for \( t \in (0, \infty) \) and \( s \in [0, t) \). However, it is easy to see that if \( X \) has an exponential distribution, then (10.1) holds for all \( t \in (0, \infty) \) and \( s \in [0, \infty) \). Thus, our converse is best stated as follows:
Theorem 10.2. Suppose that (10.1) holds for \( s = 0 \) and for \( s = t \), for all \( t \in (0, \infty) \). Then \( X \) has an exponential distribution.

Proof. The hypotheses are that

\[
\begin{align*}
  f(0) &= [1 - F(t)] \sum_{n=0}^{\infty} f(nt), \quad t \in (0, \infty) \\
  f(t) &= [1 - F(t)] \sum_{n=0}^{\infty} f([n+1]t), \quad t \in (0, \infty)
\end{align*}
\]

(10.2)

(10.3)

Combining (10.2) and (10.3) with a bit of algebra gives \( f(t) = f(0)F(t) \) for \( t \in [0, \infty) \) and hence \( X \) has a constant rate distribution on \( (0, \infty) \) with rate \( f(0) \). Equivalently, \( X \) has an exponential distribution on \( (0, \infty) \) with rate \( f(0) \).

The quotient space here can also be viewed as a lexicographic product. That is, \( (0, \infty) \) is isomorphic to the lexicographic product of \( (tN, <) \) with \( ([0, t], \leq) \), as explored in Example 8.14 (with \( t = 1 \)).

10.2 Spaces isomorphic to the standard space

Let \( S \) be an interval of real numbers of the form \( [a, b] \) where \( -\infty < a < b \leq \infty \) or of the form \( (a, b] \) where \( -\infty \leq a < b < \infty \). The \( \sigma \)-algebra \( \mathcal{S} \) is the usual Borel \( \sigma \)-algebra. Let \( \Phi \) be a homeomorphism from \( S \) onto \( [0, \infty) \). Define the operator \( * \) on \( S \) by

\[
x * y = \Phi^{-1}[\Phi(x) + \Phi(y)], \quad x, y \in S
\]

then \( (S, *) \) is a positive semigroup isomorphic to \( ([0, \infty), +) \) and \( \Phi \) is an isomorphism. If \( S = [a, b] \) then \( \Phi \) is strictly increasing so \( a \) is the identity and the associated relation \( \preceq \) is the ordinary order \( \leq \). If \( S = (a, b] \) then \( \Phi \) is strictly decreasing to \( b \) is the identity and the associated relation \( \preceq \) is \( \geq \), the reverse of the standard order. If \( x, y \in S \) and \( x \preceq y \) then

\[
x^{-1} * y = \Phi^{-1}[\Phi(y) - \Phi(x)]
\]

The rule for exponentiation under \( * \) is \( x^*c = \Phi^{-1}[c\Phi(x)] \) for \( x \in S \) and \( c \in [0, \infty) \).

A left-invariant measure \( \mu \) for \( (S, *) \) (unique up to multiplication by positive constants) is given by

\[
\mu(A) = \lambda[\Phi(A)], \quad A \in \mathcal{S}
\]

where \( \lambda \) is Lebesgue measure on \( [0, \infty) \). So the underlying measure space in this chapter is \( (S, \mathcal{S}, \mu) \). If \( \Phi \) is a smooth function with derivative \( \phi \) then

\[
d\mu(x) = |\phi(x)| \, dx
\]

Because the semigroups are isomorphic, a random variable \( X \) taking values in \( S \) is exponential for the positive semigroup \( (S, *) \) if and only if \( \Phi(X) \) is exponential for the standard, positive semigroup \( ([0, \infty), +) \) (that is, \( \Phi(X) \) has an exponential distribution in the ordinary sense). In particular, a distribution will be exponential for \( (S, *) \) if and only if it is memoryless for \( (S, *) \), if and only if it has constant rate for \( (S, \preceq) \) (with respect to \( \mu \)). It therefore follows that the exponential distribution for \( (S, *) \) which has constant rate \( \alpha \in (0, \infty) \) is the distribution with right probability function \( F \) defined by

\[
F(x) = \exp[-\alpha \Phi(x)], \quad x \in S
\]

The probability density function \( f \) of \( X \) relative to the left-invariant measure \( \mu \), is of course,

\[
f(x) = \alpha \exp[-\alpha \Phi(x)], \quad x \in S
\]

by the constant rate property. The probability density function \( g \) of \( X \) relative to Lebesgue measure \( \lambda \) is given by

\[
g(x) = \alpha \exp[-\alpha \Phi(x)] |\phi(x)|, \quad x \in S
\]

In particular, note that \( \alpha |\phi| \) is the failure right rate function of \( X \) in the ordinary sense.
Suppose that $X$ is a random variable with a continuous distribution on $S$, and let $F$ denote the right probability function of $X$ for the graph $(S, \leq)$. By our usual support assumption $F(x) > 0$ for $x \in I$. Then $X$ is exponential with respect to some semigroup $(S, \cdot)$ isomorphic to $([0, \infty), +)$. Specifically, let $\Phi(x) = -\log F(x)$ so that if $S = [a, b]$, then

$$\Phi(x) = -\log[P(X \geq x)], \quad x \in [a, b]$$

and if $S = (a, b]$, then

$$\Phi(x) = -\log[P(X \leq x)], \quad x \in (a, b]$$

This is essentially a restatement of the fact that $Y = \Phi(X)$ has a standard exponential distribution. However the semigroup formulation provides some additional insights.

Let $X$ have the exponential distribution on $(S, \cdot)$ with rate parameter $\alpha \in (0, \infty)$ and right probability function $F$ given above. From Theorem 4.12, $X$ maximizes entropy over all random variables with $E[\Phi(Y)] = E[\Phi(X)] = 1/\alpha$; the maximum entropy is $1 - \ln \alpha$. Our next result deals with minimums of exponential distributions.

**Theorem 10.3.** Consider the case $S = [a, b]$ as above. Suppose that $X$ and $Y$ are independent variables that are exponential on $(S, \cdot)$ with rates $\alpha, \beta \in (0, \infty)$, respectively. Then $X \wedge Y$ is exponential for $(S, \cdot)$ and has failure rate $\alpha + \beta$. If $\alpha = \Phi(a), \beta = \Phi(b)$ then $\alpha + \beta = \Phi(a \cdot b)$.

**Proof.** The variables $X$ and $Y$ have right probability functions $F$ and $G$ for $(S, \leq)$ given by

$$F(x) = \exp[-\alpha \Phi(x)], \quad G(x) = \exp[-\beta \Phi(x)]; \quad x \in S$$

Hence $X \wedge Y$ has right probability function $FG$ which can be written

$$F(x)G(x) = \exp[-(\alpha + \beta) \Phi(x)], \quad x \in I$$

and hence $X \wedge Y$ is exponential for $(S, \cdot)$ with rate $\alpha + \beta$. The last statement follows from the definition of $\cdot$. \qed

Suppose now that $X = (X_1, X_2, \ldots)$ is the random walk on $(S, \cdot)$ associated with the exponential distribution that has rate $\alpha \in (0, \infty)$. Then $(\Phi(X_1), \Phi(X_2), \ldots)$ is the random walk on $([0, \infty), +)$ associated with the ordinary exponential distribution that has rate $\alpha$. In particular, for $n \in \mathbb{N}_+$, $\Phi(X_n)$ has the ordinary gamma distribution on $[0, \infty), +)$ with parameters $\alpha$ and $n$. It follows from the usual change of variables formula that the density $g_n$ of $Y_n$ with respect to Lebesgue measure $\lambda$ is

$$g_n(x) = \frac{\alpha^n \Phi^{n-1}(x)}{(n-1)!} \exp[-\alpha \Phi(x)]\phi(x), \quad x \in S$$

Hence, the density function $f_n$ of $X_n$ with respect to the left-invariant measure $\mu$ is

$$f_n(x) = \frac{\alpha^n \Phi^{n-1}(x)}{(n-1)!} \exp[-\alpha \Phi(x)], \quad x \in S$$

It follows that the left walk function $\gamma_n$ of order $n \in \mathbb{N}$ for $(S, \preceq)$ is

$$\gamma_n(x) = \frac{\Phi^n(x)}{n!}, \quad x \in S$$

Of course, we could also derive this last result directly.

Our characterization of exponential distributions based on independent, identically distributed variables goes like this: if $X$ and $Y$ are independent and identically distributed on $S$, then the common distribution is exponential if and only if the conditional distribution of $X$ given $X \cdot Y = z$ is uniform on $[a, z]$ if $S = [a, b)$ or uniform on $[z, b]$ if $S = (a, b]$.

In the next several subsections we explore a number of specific examples.
10.2.1 The shifted space

Our first example is a rather trivial modification of the standard space, but still leads to some helpful insights. Let \( S = [a, \infty) \) where \( a \in (0, \infty) \), and define \( \Phi : S \to [0, \infty) \) by \( \Phi(x) = x - a \) for \( x \in S \). Then \( \Phi \) is a homeomorphism from \([a, \infty)\) onto \([0, \infty)\) with inverse given by \( \Phi^{-1}(t) = t + a \) for \( t \in [0, \infty) \). Of course \( \varphi = \Phi^{-1} = 1 \). The corresponding operator \( * \) on \([a, \infty)\) is given by

\[
x * y = x + y - a, \quad x, y \in [a, \infty)
\]

The corresponding relation is the ordinary order \( \leq \) and the corresponding measure is just the original Lebesgue measure since

\[
\lambda(A - a) = \lambda(A), \quad A \in \mathcal{F}
\]

The left path function \( \gamma_n \) of order \( n \in \mathbb{N} \) is given by

\[
\gamma_n(x) = \frac{(x - a)^n}{n!}, \quad x \in [a, \infty)
\]

and the generating function \( \Gamma \) is given by

\[
\Gamma(x, t) = e^{t(x-a)}, \quad x \in [a, \infty), \quad t \in \mathbb{R}
\]

Suppose now that \( X \) is a random variable with values in \([a, \infty)\). The right probability function \( F \) of \( X \) for \(([a, \infty), \leq) \) is just the standard one, so that \( F(x) = \mathbb{P}(X \geq x) \) for \( x \in [a, \infty) \). From the general theory, \( X \) has a constant rate distribution for \(([a, \infty), \leq) \) if and only if \( X \) has a memoryless distribution for \(([a, \infty), *) \) if and only if \( X \) has an exponential distribution for \(([a, \infty), *) \) if and only if \( X - a \) has an ordinary exponential distribution on \([0, \infty)\). Hence, the distribution that has constant rate \( \beta \in (0, \infty) \) for \(([a, \infty), \leq) \) is the standard shifted exponential distribution with density \( f \) and with right probability function \( F \) for \(([a, \infty), \leq) \) given by

\[
f(x) = \beta e^{-\beta(x-a)}, \quad F(x) = e^{-\beta(x-a)}, \quad x \in [a, \infty)
\]

If \( X = (X_1, X_2, \ldots) \) is the random walk on \(([a, \infty), \leq) \) (or equivalently \(([a, \infty), *) \) corresponding to \( f \) then \( X_n \) has density function \( f_n \) given by

\[
f_n(x) = \beta^n \gamma_{n-1}(x) F(x) = \beta^n \frac{(x-a)^{n-1}}{(n-1)!} e^{-\beta(x-a)}, \quad x \in [a, \infty)
\]

Of course \( f_n \) is the shifted version of the standard gamma distribution with parameters \( n \) and \( \beta \).

But we can also consider \([a, \infty)\) with the standard addition operator \( + \). So \(([a, \infty), +) \) is a strict positive semigroup, and the associated strict partial order \( < \) is given by \( x < y \) if and only if \( x + a \leq y \). Once again we use Lebesgue measure \( \lambda \) as the invariant reference measure. Now we are outside of the setting of Section 7, but it’s interesting to compare the two spaces.

**Proposition 10.1.** The left walk function \( \gamma_n \) of order \( n \in \mathbb{N} \) for \(([a, \infty), <) \) is given by \( \gamma_n(x) = 0 \) for \( x < (n+1)a \) and

\[
\gamma_n(x) = \frac{|x - (n+1)a|^n}{n!}, \quad x \geq (n+1)a
\]

**Proof.** A “combinatorial” proof is best. Let \( n \in \mathbb{N} \) and note first that \( \gamma_n(x) = 0 \) for \( x < (n+1)a \). Suppose that \( x \geq (n+1)a \) and let \( u = (u_1, u_2, \ldots, u_n) \) satisfy

\[
(n+1)a \leq u_1 \leq u_2 \leq \cdots \leq u_n \leq x
\]

so that \((u_1, u_2, \ldots, u_n, x)\) is a path of length \( n \) in \(([a, \infty), \leq) \) terminating in \( x \). Define \( \mathbf{x} = (x_1, x_2, \ldots, x_n) \) by \( x_k = u_k - (n-k+1)a \) for \( k \in \{1, 2, \ldots, n\} \). Then

\[
x_1 < x_2 < \cdots < x_n < x
\]

so \((x_1, x_2, \ldots, x_n, x)\) is a path of length \( n \) in \(([a, \infty), <) \) terminating in \( x \). Conversely, given such a path \( \mathbf{x} \) we can recover the path \( u \) by \( u_k = x_k + (n-k+1)a \) for \( k \in \{1, 2, \ldots, n\} \). The measure of the set of paths \( \mathbf{x} \) is the same as the measure of the set of paths \( u \), which is \(|(x - (n+1)a)|^n/n!|\). 

\[\square\]
Suppose now that \( X \) is a random variable with values on \([a, \infty)\). The right probability function \( F \) of \( X \) for \(((a, \infty), \prec)\) is given by \( F(x) = P(X \geq x + a) \) for \( x \in [a, \infty)\). Clearly \( F \) does not determine the distribution of \( X \) since \( F \) gives no information about the distribution on the interval \([a, 2a)\). If \( X \) has a continuous density \( f \) then

\[
F(x) = \int_{x+a}^{\infty} f(t) \, dt, \quad x \in [a, \infty)
\]

and hence \( F'(x) = -f(x + a) \) for \( x \in [a, \infty)\).

**Proposition 10.2.** Suppose that \( X \) has the shifted exponential distribution on \([a, \infty)\) with parameter \( \beta \in (0, \infty) \). Then \( X \) has an exponential distribution for \(((a, \infty), +)\) with constant rate \( \beta e^\beta \).

**Proof.** As before, \( X \) has density function \( f \) given by \( f(x) = \beta e^{-\beta(x-n)} \) for \( x \in [a, \infty)\). The right probability function \( F \) of \( X \) for \(((a, \infty), \prec)\) is given by

\[
F(x) = P(X \geq x + a) = P(X - a \geq x) = e^{-\beta x}, \quad x \in [a, \infty)
\]

So the distribution of \( X \) is memoryless for \(((a, \infty), +)\) and has constant rate \( \beta e^\beta \).

As in the proof, let \( f \) denote the shifted exponential distribution on \([a, \infty)\) with parameter \( \beta \in (0, \infty) \), and suppose that \( X = (X_1, X_2, \ldots) \) is the random walk on \(((a, \infty), +)\) associated with \( f \). For \( n \in \mathbb{N}_+ \), \( X_n \) has density \( f_n \) given by

\[
f_n(x) = (\beta e^\beta)^n \gamma_n(x) F(x) = \beta^n \frac{(x-na)^{n-1}}{(n-1)!} e^{-\beta(x-na)}, \quad x \geq na
\]

It’s interesting to compare the two spaces. The positive semigroup \(((a, \infty), \ast)\) corresponds to the standard order \( \leq \) but has the non-standard operator \( \ast \). The strict positive semigroup \(((a, \infty), \cdot)\) has the standard operator \( + \) but has the non-standard order \( \prec \). The shifted exponential distribution on \([a, \infty)\) is an exponential distribution for both spaces.

### 10.2.2 The Beta Distribution

Let \( S = (0, 1) \) and let \( \Phi(x) = -\ln x \) for \( x \in (0, 1) \). Then \( \Phi \) is a homeomorphism from \((0, 1)\) onto \([0, \infty)\) and the associated operation \( \cdot \) is ordinary multiplication. The associated order is \( \geq \), the reverse of the ordinary order. The invariant measure \( \mu \) is given by \( d\mu(x) = (1/x) \, dx \). The left walk function \( \gamma_n \) of order \( n \in \mathbb{N} \) for \(((0, 1], \geq)\) (with respect to \( \mu \)) is given by

\[
\gamma_n(x) = (-1)^n \frac{\ln^n(x)}{n!}, \quad x \in (0, 1]
\]

The exponential distribution for \(((0, 1], \cdot)\) with constant rate \( \alpha \in (0, \infty) \) has right probability function \( F \) given by

\[
F(x) = x^\alpha, \quad x \in (0, 1]
\]

The density \( f \) with respect to \( \mu \) is given by

\[
f(x) = \alpha x^{\alpha-1}, \quad x \in (0, 1]
\]

while the density \( g \) with respect to Lebesgue measure \( \lambda \) is given by

\[
g(x) = \alpha x^{\alpha-1}, \quad x \in (0, 1]
\]

Note that this is the beta distribution with parameters \( \alpha \) and \( \pi \) and the special case \( \alpha = 1 \) gives the uniform distribution on \((0, 1] \). If \( X \) has the exponential distribution on \(((0, 1], \cdot)\) with rate parameter \( \alpha \in (0, \infty) \) then \( X \) maximizes entropy over all random variables \( Y \) taking values in \((0, 1]\) with \( E(-\ln Y) = 1/\alpha \).

Suppose that \( X = (X_1, X_2, \ldots) \) is the random walk on \(((0, 1], \cdot)\) associated with the exponential distribution on \(((0, 1], \cdot)\) that has rate \( \alpha \in (0, \infty) \). For \( n \in \mathbb{N}_+ \), \( X_n \) has density function \( f_n \) given by

\[
f_n(x) = \alpha^n (-1)^{n-1} \frac{\ln^{n-1}(x)}{(n-1)!} x^\alpha, \quad x \in (0, 1]
\]

with respect to \( \mu \), and density \( g_n \) given by

\[
g_n(x) = \alpha^n (-1)^{n-1} \frac{\ln^{n-1}(x)}{(n-1)!} x^{\alpha-1}, \quad x \in (0, 1]
\]

with respect to Lebesgue measure \( \lambda \).
10.2.3 The Pareto Distribution

Let $S = [1, \infty)$ and let $\Phi(x) = \ln x$ for $x \in [1, \infty)$. Then $\Phi$ is a homeomorphism from $[1, \infty)$ onto $[0, \infty)$ and the associated operation $\cdot$ is ordinary multiplication. The associated order is the ordinary order $\leq$. The invariant measure $\mu$ is given by $d\mu(x) = (1/x) \, dx$. The left walk function $\gamma_n$ of order $n \in \mathbb{N}$ for $([1, \infty), \leq)$ (with respect to $\mu$) is given by

$$
\gamma_n(x) = \frac{\ln^n(x)}{n!}, \quad x \in [1, \infty)
$$

The exponential distribution with constant rate $\alpha \in (0, \infty)$ for $([1, \infty), \cdot)$ has right probability function $F$ given by

$$
F(x) = x^{-\alpha}, \quad x \in [1, \infty)
$$

The density function $f$ with respect to $\mu$ is given by

$$
f(x) = \alpha x^{-\alpha}, \quad x \in [1, \infty)
$$

while the density $g$ with respect to Lebesgue measure $\lambda$ is given by

$$
g(x) = \alpha x^{-(\alpha+1)}, \quad x \in [1, \infty)
$$

Note that this is the Pareto distribution with shape parameter $\alpha$. This distribution maximizes entropy over all random variables $Y$ taking values in $[1, \infty)$ with $\mathbb{E}(\ln Y) = 1/\alpha$.

Suppose that $X = (X_1, X_2, \ldots)$ is the random walk on $([1, \infty), \cdot)$ corresponding to the exponential distribution with parameter $\alpha \in (0, \infty)$. For $n \in \mathbb{N}_+$, $X_n$ has density function $f_n$ given by

$$
f_n(x) = \alpha^n \frac{\ln^{n-1}(x)}{(n-1)!} x^{\alpha}, \quad x \in [1, \infty)
$$

with respect to $\mu$, and density $g_n$ given by

$$
g_n(x) = \alpha^n \frac{\ln^{n-1}(x)}{(n-1)!} x^{\alpha-1}, \quad x \in [1, \infty)
$$

with respect to Lebesgue measure $\lambda$.

10.2.4 An application to Brownian functionals

Let $S = [0, 1/2)$ and let $\Phi(x) = x/(1 - 2x)$ for $x \in [0, 1/2)$. Then $\Phi$ is a homeomorphism from $[0, 1/2)$ onto $[0, \infty)$ and the corresponding semigroup operation $\ast$ is given by

$$
x \ast y = \frac{x + y - 4xy}{1 - 4xy}; \quad x, y \in [0, 1/2)
$$

The invariant measure $\mu$ is given by

$$
d\mu(x) = \frac{1}{(1 - 2x)^2} \, dx
$$

The left walk function $\gamma_n$ of order $n \in \mathbb{N}$ for $([0, 1/2), \leq)$ (with respect to $\mu$) is given by

$$
\gamma_n(x) = \frac{1}{n!} \left( \frac{x}{1 - 2x} \right)^n, \quad x \in [0, 1/2)
$$

and the associated relation is the ordinary order $\leq$. The positive semigroup $([0, 1/2), \ast)$ occurs in the study of generalized Brownian functionals [20].

The exponential distribution on $([0, 1/2), \ast)$ with rate $\alpha \in (0, \infty)$ has right probability function $F$ given by

$$
F(x) = \exp \left( -\alpha \frac{x}{1 - 2x} \right), \quad x \in [0, 1/2)
$$

The density function $f$ with respect to $\mu$ is given by

$$
f(x) = \alpha \exp \left( -\alpha \frac{x}{1 - 2x} \right), \quad x \in [0, 1/2)
$$
while the density function \( g \) with respect to Lebesgue measure \( \lambda \) is given by
\[
g(x) = \alpha \frac{1}{(1-2x)^2} \exp\left(-\alpha \frac{x}{1-2x}\right), \quad x \in [0, 1/2)
\]
This distribution maximizes entropy over all random variables \( Y \) in \([0, 1/2)\) with \( \mathbb{E}[Y/(1-2Y)] = 1/\alpha \).

Suppose that \( X = (X_1, X_2, \ldots) \) is the random walk on \((0, 1/2, \ast)\) associated with the exponential distribution that has parameter \( \alpha \). For \( x \in [0, 1/2) \), \( X \) has density \( f_n \) given by
\[
f_n(x) = \frac{\alpha^n}{(n-1)!} \left( \frac{x}{1-2x} \right)^{n-1} \exp\left(-\alpha \frac{x}{1-2x}\right), \quad x \in [0, 1/2)
\]
with respect to \( \mu \) and has density \( g_n \) given by
\[
g_n(x) = \frac{\alpha^n}{(n-1)!} \left( \frac{x}{1-2x} \right)^{n-1} \exp\left(-\alpha \frac{x}{1-2x}\right), \quad x \in [0, 1/2)
\]
with respect to Lebesgue measure \( \lambda \).

### 10.3 Applications to reliability

The positive semigroup formulation provides a way to measure the relative aging of one lifetime distribution relative to another. In this section, we fix an interval \([a, b)\) with \(-\infty < a < b \leq \infty\), and we consider only random variables with continuous distributions on \([a, b)\). If \( X \) is such a random variable then the right probability function \( F \) of \( X \) for the graph \(([a, b), \leq)\) has the form
\[
F(x) = \mathbb{P}(X \geq x) = e^{-R(x)}, \quad x \in [a, b)
\]
where \( R \) is a homeomorphism from \([a, b)\) onto \([0, \infty)\). If \( X \) is interpreted as a random lifetime, then \( F \) is called the reliability function and \( R \) is called the cumulative failure right rate function. (For a basic introduction to reliability, see [10].) As noted in the last section, if we define
\[
x \ast y = R^{-1}[R(x) + R(y)], \quad x, y \in [a, b)
\]
then \(([a, b) \ast)\) is a positive semigroup isomorphic to \(([0, \infty), +)\), and \( X \) has an exponential distribution on \((a, b) \ast)\). Note that the graph associated with \(([a, b), \ast)\) is in fact \((a, b), \leq)\) since \( R \) is strictly increasing.

Suppose now that \( X \) and \( Y \) are random variables on \([a, b)\) with cumulative failure right rate functions \( R \) and \( S \), and semigroup operations \( \bullet \) and \( \ast \), respectively. A natural way to study the relative aging of \( X \) relative to \( Y \) is to study the aging of \( X \) on \(([a, b), \ast)\). Note that the cumulative failure right rate function and the reliability function of \( X \) are still \( R \) and \( e^{-R} \), respectively, when considered on \((a, b), \ast)\), because the associated graph is just \((a, b), \leq)\). Thus, these functions are invariants.

**Definition 10.1.** Random variable \( X \) is exponential relative to \( Y \) if \( X \) has an exponential distribution on \((a, b), \ast)\).

**Theorem 10.4.** \( X \) is exponential relative to \( Y \) if and only if the semigroup operators \( \bullet \) and \( \ast \) are the same. The exponential relation defines an equivalence relation on the collection of continuous distributions on \([a, b)\).

**Proof.** Note that \( X \) is exponential relative to \( Y \) if and only if \( S = cR \) for some positive constant \( c \).

Next we consider the increasing failure rate property, the strongest of the basic aging properties. If \( R \) and \( S \) have positive derivatives on \([a, b)\), then the density function \( f \) of \( X \) relative to the invariant measure on \((a, b), \ast)\) is given by
\[
f(x) = e^{-R(x)} \frac{R'(x)}{S'(x)}, \quad x \in [a, b)
\]
Thus, the failure right rate function of \( X \) on \((a, b), \ast)\) is obtained by dividing the density function by the reliability function, and hence is \( R'/S' \).

**Lemma 10.1.** \( R'/S' \) is increasing on \([a, b)\) if and only if \( R \) is convex on \((a, b), \ast)\):
\[
R(x \ast h) - R(x) \leq R(y \ast h) - R(y), \quad \text{for every } x, y, h \in [a, b) \text{ with } x \leq y
\]
Thus, we will use the convexity condition for our definitions, since it is more general by not requiring that $R$ and $S$ be differentiable.

**Definition 10.2.** Random variable $X$ has increasing failure rate (IFR) relative to $Y$ if $R$ is convex on $([a,b],*)$, and $X$ has decreasing failure rate (DFR) relative to $Y$ if $R$ is concave on $([a,b],*)$.

**Theorem 10.5.** $X$ has increasing failure rate relative to $Y$ if and only if $Y$ has decreasing failure rate relative to $X$ if and only if $$(x \ast h) \cdot y \leq (y \ast h) \cdot x \text{ for all } x, y, h \in [a,b) \text{ with } x \leq y$$

The IFR relation defines a partial order on the collection of continuous distributions on $[a,b)$, modulo the exponential equivalence in Theorem 10.4.

**Proof.** Note that $X$ has increasing failure rate relative to $Y$ if and only if the distribution on $[0,\infty)$ with cumulative right rate function $R \circ S^{-1}$ is IFR in the ordinary sense.

Next, for $x \in (a,b)$, the failure rate average over $[a,x)$ is the cumulative failure rate over $[a,x)$ divided by the length of this interval, as measured by the invariant measure on $([a,b],*)$. This length is simply $S(x)$, and hence the average failure rate function of $X$ on $([a,b],*)$ is $R/S$.

**Definition 10.3.** $X$ has increasing failure rate average (IFRA) relative to $Y$ if $R/S$ is increasing on $[a,b)$ and decreasing failure rate average (DFRA) relative to $Y$ if $R/S$ is decreasing on $[a,b)$.

**Theorem 10.6.** $X$ has increasing failure rate average relative to $Y$ if and only if $Y$ has decreasing failure rate average relative to $X$ if and only if $$x^{\ast \alpha} \leq x^{\ast \alpha} \text{ for all } x \in [a,b) \text{ and } \alpha \geq 1$$

The IFRA relation defines a partial order on the collection of continuous distributions on $[a,b)$, modulo the exponential equivalence in Theorem 10.4.

**Proof.** Note that $X$ has increasing failure rate relative to $Y$ if and only if the distribution on $[0,\infty)$ with cumulative right rate function $R \circ S^{-1}$ is IFRA in the ordinary sense.

Next, $X$ is new better than used relative to $Y$ if the conditional reliability function of $x^{\ast(-1)} \ast X$ given $X \geq x$ is dominated by the reliability function of $X$. Here is an equivalent definition:

**Definition 10.4.** $X$ is new better than used (NBU) relative to $Y$ if $$P(X \geq x \ast y) \leq P(X \geq x)P(X \geq y), \quad x, y \in [a,b)$$ $X$ is new worse than used (NWU) relative to $Y$ if $$P(X \geq x \ast y) \geq P(X \geq x)P(X \geq y), \quad x, y \in [a,b)$$

**Theorem 10.7.** $X$ is new better than used relative to $Y$ if and only if $Y$ is new worst than used relative to $X$ if and only if $$x \cdot y \leq x \ast y \text{ for } x, y \in [a,b)$$

The NBU relation defines a partial order on the collection of continuous distributions on $I$, modulo the exponential equivalence in Theorem 10.4.

**Proof.** $X$ is new better than used relative to $Y$ if and only if the distribution on $[0,\infty)$ with cumulative right rate function $R \circ S^{-1}$ is NBU in the ordinary sense.

**Theorem 10.8.** The relative aging properties for distributions on $[a,b)$ are related as follows:

$$\text{IFR} \Rightarrow \text{IFRA} \Rightarrow \text{NBU}$$

Equivalently, the NBU partial order extends the IFRA partial order, which in turn extends the IFR partial order.

**Proof.** It’s well known that the standard aging properties on $[0,\infty)$ are related as stated.
10.3.1 Examples

In the following examples, we consider several two-parameter families of distributions. In each case, \( \alpha \) is the exponential parameter while \( \beta \) is an “aging parameter” that determines the relative aging.

**Example 10.1** (Weibull distribution). Let \( S = [0, \infty) \) and for fixed \( \beta \in (0, \infty) \), let \( R \) be given by \( R(x) = x^\beta \) for \( x \in [0, \infty) \). The corresponding semigroup operator \( * \) is given by

\[
x * y = (x^\beta + y^\beta)^{1/\beta}, \quad x, y \in [0, \infty)
\]

The exponential distribution on \( (0, \infty), * \) with rate \( \alpha \in (0, \infty) \) has reliability function \( F \) given by

\[
F(x) = \exp(-\alpha x^\beta), \quad x \in [0, \infty)
\]

Of course, this is a Weibull distribution with shape parameter \( \beta \). In the usual formulation, the rate parameter \( \alpha \) is written as \( \alpha = R(c) = c^\beta \) where \( c \in (0, \infty) \) is the scale parameter. In the ordinary sense, the Weibull distribution has decreasing failure rate if \( 0 < \beta < 1 \), is exponential if \( \beta = 1 \), and has increasing failure rate if \( \beta > 1 \). A stronger statement, from Theorem 10.5 is that the Weibull distribution with shape parameter \( \beta_1 \in (0, \infty) \) is IFR relative to the Weibull distribution with shape parameter \( \beta_2 \in (0, \infty) \) if and only if \( \beta_1 \leq \beta_2 \).

**Example 10.2** (Extreme value distribution). Again let \( S = [0, \infty) \) and for fixed \( \beta \in (0, \infty) \), let \( R \) be given by \( R(x) = e^{\beta x} - 1 \) for \( x \in [0, \infty) \). The corresponding semigroup operator \( * \) is given by

\[
x * y = \frac{1}{\beta} \ln(e^{\beta x} + e^{\beta y} - 1), \quad x, y \in [0, \infty)
\]

The exponential distribution on \( (0, \infty), * \) with rate \( \alpha \in (0, \infty) \) has reliability function \( F \) given by

\[
F(x) = \exp[-\alpha(e^{\beta x} - 1)], \quad x \in [0, \infty)
\]

This is the modified extreme value distribution with parameters \( \alpha \) and \( \beta \). In the ordinary sense, these distributions are IFR for all parameter values. On the other hand, a distribution with parameter \( \beta_1 \in (0, \infty) \) is IFR with respect to a distribution with parameter \( \beta_2 \in (0, \infty) \) if and only if \( \beta_1 \leq \beta_2 \).

**Example 10.3** (Pareto distribution). Again let \( S = [0, \infty) \) and for fixed \( \beta \in (0, \infty) \), let \( R \) be given by \( R(x) = \ln(x + \beta) - \ln(\beta) \) for \( x \in [0, \infty) \). The corresponding semigroup operator \( * \) is given by

\[
x * y = x + y + \frac{xy}{\beta}, \quad x, y \in [0, \infty)
\]

The exponential distribution on \( (0, \infty), * \) with rate \( \alpha \in (0, \infty) \) has reliability function \( F \) given by

\[
F(x) = \left( \frac{\beta}{x + \beta} \right)^\alpha, \quad x \in [0, \infty)
\]

This is a two-parameter family of Pareto distribution. In the ordinary sense, these distributions are DFR for all parameter values. On the other hand, a distribution with parameter \( \beta_1 \in (0, \infty) \) is DFR with respect to a distribution with parameter \( \beta_2 \in (0, \infty) \) if and only if \( \beta_1 \leq \beta_2 \).

**Example 10.4** (Beta distribution). Let \( S = [0, 1) \) and for fixed \( \beta \in (0, \infty) \), let \( R \) be given by \( R(x) = -\ln(1 - x^{-\beta}) \) for \( x \in [0, 1) \). The corresponding semigroup operator \( * \) is given by

\[
x * y = (x^\beta + y^\beta - x^\beta y^\beta)^{1/\beta}; \quad x, y \in [0, 1)
\]

The exponential distribution on \( (0, 1), * \) with rate \( \alpha \in (0, \infty) \) has reliability function \( F \) given by

\[
F(x) = (1 - x^\beta)^\alpha, \quad x \in [0, 1)
\]

Note that if \( \alpha = 1 \) or \( \beta = 1 \), the distribution is beta; if \( \alpha = \beta = 1 \), the distribution is uniform. In the ordinary sense, these distributions are IFR if \( \beta \geq 1 \), but are neither NBU nor NWU if \( 0 < \beta < 1 \). On the other hand, a distribution with parameter \( \beta_1 \in (0, \infty) \) is DFR with respect to a distribution with parameter \( \beta_2 \in (0, \infty) \) if and only if \( \beta_1 \leq \beta_2 \).
10.3.2 Strong aging properties

Consider an “aging property”, and the corresponding “improvement property”, for continuous distributions in the standard space \((0, \infty, +)\). We are interested in characterizing those aging properties that can be extended to relative aging properties for continuous distributions on an arbitrary interval \([a, b]\). Moreover, we want the relative aging property to define a partial order on the distributions, modulo the exponential equivalence in Theorem 10.4, just as the IFR, IFRA, and NBU properties do. Such a characterization would seem to describe “strong” aging properties.

**Definition 10.5.** A **strong aging property** satisfies the following conditions:

(a) A distribution both ages and improves if and only if the distribution is exponential.

(b) The distribution with cumulative rate \(R\) ages if and only if the distribution with cumulative rate \(R^{-1}\) improves.

(c) If the distributions with cumulative rates \(R\) and \(S\) age, then the distribution with cumulative rate \(R \circ S\) ages.

(d) If a distribution is IFR then the distribution ages.

The last condition is to ensure that the property does capture some idea of aging, and to incorporate the fact that the IFR condition is presumably the strongest aging property.

**Definition 10.6.** Consider a strong aging property, and suppose that \(X\) and \(Y\) are random variables with continuous distributions on \([a, b]\) having cumulative rate functions \(R\) and \(S\), respectively. We say that \(X\) ages (improves) relative to \(Y\) if the distribution on \([0, \infty)\) with cumulative right rate function \(R \circ S^{-1}\) ages (improves), respectively.

**Theorem 10.9.** Consider a strong aging property. The corresponding aging relation defines a partial order on the equivalence class of continuous distributions on \([a, b]\), modulo the exponential equivalence.

**Proof.** The proof is straightforward.

**Theorem 10.10.** The IFR, IFRA, and NBU properties are strong aging properties.

**Proof.** This follows from our previous theorems on IFR, IFRA, and NBU.

From our point of view, the conditions in Definition 10.5 are natural for a strong aging property. Conditions (a) and (d) are often used in the literature to justify aging properties, but Conditions (b) and (c) seem to have been overlooked, even though they are essential for the partial order result in the theorem.

Not all of the common aging properties are strong. A random variable \(X\) with a continuous distribution on \([0, \infty)\) is **new better than used in expectation** (NBUE) if

\[
E(X - t | X \geq t) \leq E(X), \quad t \in [0, \infty)
\]

and of course is **new worse than used in expectation** (NWUE) if the inequality is reversed.

**Theorem 10.11.** NBUE is not a strong aging property.

**Proof.** Define \(R : [0, \infty) \rightarrow [0, \infty)\) by

\[
R(t) = \begin{cases} 
  at, & 0 \leq t < 1 \\
  a + b(t - 1), & 1 \leq t < 2 \\
  a + b(t - 1) + c(t - 2), & t \geq 2 
\end{cases}
\]

Positive constants \(a, b,\) and \(c\) can be chosen such that the distribution with cumulative right rate function \(R\) is NBUE, but the distribution with cumulative right rate function \(R^{-1}\) is not NWUE.
10.4 Norm Graphs

In this section we study an interesting collection of graphs on \( \mathbb{R}^n \) that are induced (in the sense of Chapter 7) by the standard graph \((0, \infty), \leq\) and the norm functions. For \( n \in \mathbb{N}_+ \), we give \( \mathbb{R}^n \) the usual Borel \( \sigma \)-algebra \( \mathcal{B}^n \) and \( n \)-dimensional Lebesgue measure \( \lambda^n \) as the reference measure. For \( m \in \mathbb{N} \) and \( m \leq n \), we also use \( \lambda^m \) to denote Lebesgue measure (or more technically Hausdorff measure) on an \( m \)-dimensional Riemannian (smooth) manifold of \( \mathbb{R}^n \). In particular, recall that \( \lambda^0 \) is counting measure. For \( k \in [1, \infty) \), recall that the \( k \)-norm on \( \mathbb{R}^n \) is given by

\[
\| x \|_k = \left( \sum_{i=1}^n |x_i|^k \right)^{1/k}, \quad x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n
\]

The \( \infty \) norm is defined by

\[
\| x \|_\infty = \max\{|x_i| : i \in \{1, 2, \ldots, n\} \}, \quad x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n
\]

When \( n = 1 \), the norms are all the same: \( \| x \|_k = |x| \) for \( k \in [1, \infty] \) and \( x \in \mathbb{R} \). When \( n \in \{2, 3, \ldots\} \), the norms are all different, of course. For \( n \in \mathbb{N}_+ \) and \( k \in [1, \infty] \), the mapping \( \| \cdot \|_k \) is a measurable function from \( \mathbb{R}^n \) onto \((0, \infty).

**Definition 10.7.** For \( n \in \mathbb{N}_+ \) and \( k \in [1, \infty) \), let \((\mathbb{R}^n, \Rightarrow_k)\) denote the graph induced by \( \| \cdot \|_k \) and the standard graph \((0, \infty), \leq\), so that \( x \Rightarrow_k y \) if and only if \( \| x \|_k \leq \| y \|_k \) for \( x, y \in \mathbb{R}^n \).

Note that \( \Rightarrow_k \) is reflexive and transitive.

**Definition 10.8.** For \( n \in \mathbb{N}_+ \) and \( k \in [1, \infty) \), define

(a) \( b_{n,k} = \lambda^n \{ x \in \mathbb{R}^n : \| x \|_k \leq 1 \} \), the volume of the unit ball in \( \mathbb{R}^n \) under the \( k \)-norm.

(b) \( c_{n,k} = \lambda^{n-1} \{ x \in \mathbb{R}^n : \| x \|_k = 1 \} \), the surface area of the unit ball in \( \mathbb{R}^n \) under the \( k \)-norm.

It’s well known [37] that

\[
b_{n,k} = \frac{2^n \Gamma(1 + 1/k)^n}{\Gamma(1 + n/k)}, \quad n \in \mathbb{N}_+, \ k \in [1, \infty)
\]

and that \( b_{n,\infty} = 2^n \) for \( n \in \mathbb{N}_+ \), the limiting value of \( b_{n,k} \) as \( k \to \infty \). For \( n \in \mathbb{N}_+ \), \( k \in [1, \infty] \), and \( t \in (0, \infty) \),

\[
b_{n,k} t^n = \lambda^n \{ x \in \mathbb{R}^n : \| x \|_k \leq t \}
\]

\[
c_{n,k} t^{n-1} = \lambda^{n-1} \{ x \in \mathbb{R}^n : \| x \|_k = t \}
\]

So \( b_{n,k} t^n \) is the volume of the ball of radius \( t \) and \( c_{n,k} t^{n-1} \) is the surface area of the ball of radius \( t \), both relative to the \( k \)-norm. As functions of \( t \in [0, \infty) \), the surface area is a multiple of the derivative of the volume, or equivalently, for \( n \in \mathbb{N}_+ \) and \( k \in [1, \infty] \) there exists \( a_{n,k} \in (0, \infty) \) such that

\[
c_{n,k} = a_{n,k} b_{n,k}
\]

For most values of \( n \in \mathbb{N}_+ \) and \( k \in [1, \infty] \), there is no simple closed formula for \( a_{n,k} \), but there are a few exceptions:

- \( a_{n,2} = 1 \) for \( n \in \mathbb{N}_+ \) so

\[
c_{n,2} = nb_{n,2} = 2 \frac{\pi^{n/2}}{\Gamma(n/2)}, \quad n \in \mathbb{N}_+
\]

- \( a_{n,\infty} = 1 \) for \( n \in \mathbb{N}_+ \) so

\[
c_{n,\infty} = nb_{n,\infty} = 2^n, \quad n \in \mathbb{N}_+
\]

- \( a_{n,1} = \sqrt{n} \) so

\[
c_{n,1} = n \sqrt{b_{n,1}} = \frac{2^n \sqrt{n}}{(n-1)!}, \quad n \in \mathbb{N}_+
\]
The basic Assumption 7.1 is satisfied:
\[ \lambda^n(A) = \int_0^\infty \frac{1}{a_{n,k}} \lambda^{n-1}(A \cap B_k^n(t))dt, \quad A \in \mathbb{R}^n \]
where \( \mathbb{R}^n_k = \{ x \in \mathbb{R}^n : \|x\|_k = t \} \), the surface of the ball of radius \( t \in [0, \infty) \) relative to the \( k \) norm, and the inverse image at \( t \) of the mapping \( x \mapsto \|x\|_k \) from \( \mathbb{R}^n \) onto \( [0, \infty) \). This is essentially the co-area formula for computing \( \lambda^n(A) \). Again in the notation Chapter 7, note also that for \( n \in \mathbb{N}_+ \) and \( k \in [1, \infty] \), the function \( \beta_{n,k} \) is given by
\[ \beta_{n,k}(t) = \frac{1}{a_{n,k}} \lambda^{n-1}(B_k^n(t)) = \frac{c_{n,k}}{a_{n,k}} t^{n-1} = nb_{n,k}t^{n-1}, \quad t \in [0, \infty) \]
Going forward, we will suppress the dependence on \( n \) and \( k \) except when necessary. Our first result is the walk function.

**Proposition 10.3.** The left walk function \( w_m \) of order \( m \in \mathbb{N}_+ \) for \((\mathbb{R}^n, \Rightarrow_k)\) is given by
\[ w_m(x) = \frac{b_{n,k}^m}{m!} \|x\|_k^{mn}, \quad x \in \mathbb{R}^n \]

**Proof.** This follows from \( w_0 = 1 \) and the recursion relation
\[ w_{m+1}(x) = \int_{y \Rightarrow x} w_m(y) d\lambda^n(y), \quad x \in \mathbb{R}^n \]
The integral can be evaluated using the co-area formula, as noted above. There is also a simple “combinatorial” argument. For \( x \in \mathbb{R}^n \), the measure of the set
\[ \{ (y_1, y_2, \ldots, y_m) : y_i \in \mathbb{R}^n, \|y_i\|_k \leq \|x\|_k \text{ for each } i \in \{1, 2, \ldots, m\} \} \]
is of course \( b_{n,k}^m \|x\|_k^{mn} \). But another algorithm for constructing this set this set is to select a walk \((x_1, x_2, \ldots, x_m)\) of length \( m \) ending in \( x \) for the graph \((\mathbb{R}^n, \Rightarrow_k)\) and then select an ordering \((y_1, y_2, \ldots, y_m)\) of \((x_1, x_2, \ldots, x_m)\). By definition, the measure of the set of walks is \( w_m(x) \), and there are \( m! \) permutations of each walk, so it follows that
\[ m! w_m(x) = b_{n,k}^m \|x\|_k^{mn} \]
\[ \square \]

Note that \( w_m(x) \) is the left walk function \( t \mapsto t^m/m! \) of order \( m \) for \(([0, \infty), \leq)\), evaluated at \( b_{n,k} \|x\|_k^n \) which is the volume of the ball of radius \( \|x\|_k \) in the \( k \) norm. We will see this pattern repeatedly. In particular, it follows that the left generating function of \((\mathbb{R}^n, \Rightarrow_k)\) is the left generating function of \(([0, \infty), \leq)\) evaluated at \( b_{n,k} \|x\|_k^n \):

**Corollary 10.1.** The left generating function \( W \) of \((\mathbb{R}^n, \Rightarrow_k)\) is given by
\[ W(x, t) = \exp (b_{n,k} \|x\|_k^n t), \quad x \in \mathbb{R}^n, \ t \in \mathbb{R} \]

**Proof.** This follows easily from the definition:
\[ W(x, t) = \sum_{m=0}^{\infty} w_m(x) t^m, \quad x \in \mathbb{R}^n, \ t \in \mathbb{R} \]
\[ \square \]

Suppose now that \( X = (X_1, X_2, \ldots, X_n) \) is a random variable in \( \mathbb{R}^n \), so that \( \|X\|_k \) is the corresponding induced variable in \([0, \infty)\). The right probability function \( \tilde{F} \) of \( \|X\|_k \) for \(([0, \infty), \leq)\) is the usual reliability function:
\[ \tilde{F}(t) = \mathbb{P} (\|X\|_k \geq t), \quad t \in [0, \infty) \]
Hence the right probability function \( F \) of \( X \) for \((\mathbb{R}^n, \Rightarrow_k)\) is
\[ F(x) = \tilde{F} (\|x\|_k), \quad x \in \mathbb{R}^n \]
If $X$ has density function $f$ then $\|X\|_k$ has density function $\hat{f}$ given by

$$\hat{f}(t) = \int_{\mathbb{R}^n_k(t)} \frac{1}{a_{n,k}} f(x) d\lambda^{n-1}(x), \quad t \in [0, \infty)$$

We assume that $\hat{f}(t) > 0$ for almost all $t \in [0, \infty)$, so that $\hat{F}(t) > 0$ for almost all $t \in [0, \infty)$ and $F(x) > 0$ for almost all $x \in \mathbb{R}^n$. For $t \in [0, \infty)$, the conditional distribution of $X$ given $\|X\|_k = t$ has density function $x \mapsto f(x \mid t) = f(x) / \hat{f}(t)$ on $\mathbb{R}^n_k(t)$. The general results of Chapter 7 apply, of course, but as usual we are most interested in constant rate distributions. Here are the main results:

**Theorem 10.12.** For $n \in \mathbb{N}_+$ and $k \in [1, \infty]$, the graph $(\mathbb{R}^n, \Rightarrow_k)$ has a unique distribution with constant rate $\alpha$ for each $\alpha \in (0, \infty)$. For this distribution,

(a) $\|X\|_k$ has the Weibull distribution with shape parameter $n$ and scale parameter $(\alpha b_{n,k})^{-1/n}$, with density $\hat{f}$ defined by

$$\hat{f}(t) = \alpha n b_{n,k} t^{n-1} \exp(-\alpha b_{n,k} t^n), \quad t \in [0, \infty)$$

(b) $X$ has density function $f$ defined by

$$f(x) = \alpha \exp(-\alpha b_{n,k} \|x\|^n_k), \quad x \in \mathbb{R}^n$$

(c) The conditional distribution of $X$ given $\|X\|_k = t$ is uniform on $\mathbb{R}^n_k(t)$ for $t \in [0, \infty)$.

**Proof.** The results follow from Theorem 7.1. Specifically, $X$ has constant rate $\alpha \in (0, \infty)$ for $(\mathbb{R}^n, \Rightarrow_k)$ if and only if $\|X\|_k$ has rate function $\hat{r}$ for $([0, \infty), \leq)$ where $\hat{r}(t) = \alpha \beta_{n,k}(t) = \alpha n b_{n,k} t^{n-1}$ for $t \in [0, \infty)$. So $\|X\|_k$ has right probability function $\hat{F}$ for $([0, \infty), \leq)$ given by

$$\hat{F}(t) = \exp \left( - \int_0^t \hat{r}(s) \, ds \right) = \exp(-\alpha b_{n,k} t^n), \quad t \in [0, \infty)$$

which we recognize as the Weibull distribution in (a). And then $X$ has right probability function $F$ for $(\mathbb{R}^n, \Rightarrow_k)$ given by $F(x) = \hat{F}(\|x\|_k)$ for $x \in \mathbb{R}^n$, and has density function $f$ given by $f(x) = \alpha \hat{F}(\|x\|_k)$ for $x \in \mathbb{R}^n$.

Note that $f(x)$ is the ordinary exponential density function for $([0, \infty), +)$, with rate $\alpha$, evaluated at $b_{n,k} \|x\|^n_k$, which again is the volume of the ball of radius $\|x\|_k$ in the $k$ norm.

**Corollary 10.2.** Let $n \in \mathbb{N}_+$ and $k \in [1, \infty]$, and suppose that $X = (X_1, X_2, \ldots, X_n)$ has the distribution with constant rate $\alpha \in (0, \infty)$ for the graph $(\mathbb{R}^n, \Rightarrow_k)$.

(a) $X$ is an exchangeable sequence of random variables.

(b) $X$ is a sequence of identically distributed variables.

(c) The distribution of $X$ is symmetric in each variable.

(d) $X$ is a sequence of mean 0 and pairwise-uncorrelated variables.

**Proof.** These results follow from the density function in Theorem 10.12.

(a) Note that $f(x_1, x_2, \ldots, x_n)$ is invariant under a permutation of $(x_1, x_2, \ldots, x_n)$.

(b) This follows from (a).

(c) Note that if $(e_1, e_2, \ldots, e_n) \in \{-1, 1\}^n$ then $f(e_1x_1, e_2x_2, \ldots, e_nx_n) = f(x_1, x_2, \ldots, x_n)$. It follows that $(e_1X_1, e_2X_2, \ldots, e_nX_n)$ has the same distribution as $X$.

(d) This follows from (c). For distinct $i, j \in \{1, 2, \ldots, n\}$, we have $E(X_i) = E(-X_i) = -E(X_i)$ and $E(X_iX_j) = E(-X_iX_j) = -E(X_iX_j)$. Hence $E(X_i) = 0$ and then $\text{cov}(X_i, X_j) = E(X_iX_j) = 0$. 

\qed
Theorem 10.12. If \( n \) and \( g \) are as in Corollary 10.4, with parameters \( n, \alpha \in (0, \infty) \), then \( \alpha \) and \( n \) are odd and positive integers, and forms an interesting special class of generalized normal distributions.

Corollary 10.4. Let \( n \in \mathbb{N}_+ \) and suppose that \( X = (X_1, X_2, \ldots, X_n) \) has the distribution with constant rate \( \alpha \in (0, \infty) \) for the graph \((\mathbb{R}^n, \Rightarrow)\). Then \( X \) is a sequence of independent, identically distributed variables with common density function \( g \) on \( \mathbb{R} \) define by

\[
g(x) = \alpha^{1/n} \exp(-\alpha b_n |x|^n), \quad x \in \mathbb{R}
\]

Proof. These results follow immediately from part (b) of Theorem 10.12 and the factorization theorem for independence.

Chapter 10. Standard Continuous Spaces

Theorem 10.12. The distribution on \( \mathbb{R} \) defined by the density function in (b) is the generalized normal distribution (see [12] and [36]) with shape parameter \( n \) and scale parameter

\[
\frac{n}{2\alpha^{1/n} \Gamma(1/n)}
\]

The distribution actually makes sense for any \( n \in (0, \infty) \), not just positive integers, and forms an interesting special class of generalized normal distributions.

Corollary 10.5. Suppose that \( X \) has the distribution on \( \mathbb{R} \) given in (b) of Corollary 10.4, with parameters \( n, \alpha \in (0, \infty) \).

(a) If \( n = 1 \), \( X \) has the Laplace (double exponential) distribution with parameter \( 2\alpha \).

(b) If \( n = 2 \), \( X \) has the normal distribution with mean 0 and variance \( 1/2\alpha \).

(c) If \( m \in \mathbb{N}_+ \) is odd then \( \mathbb{E}(X^m) = 0 \). If \( m \in \mathbb{N}_+ \) is even then

\[
\mathbb{E}(X^m) = \alpha^{-m/2} \left( \frac{n}{2} \right)^m \frac{\Gamma(m+1)}{\Gamma(m+1) \left( \frac{1}{n} \right)}
\]

(d) The distribution of \( X \) converges to the uniform distribution on \([-\frac{1}{2}, \frac{1}{2}]\) as \( n \to \infty \).

(e) \( X \) has entropy \( H(X) = (1 + \ln \alpha)/n \).

Proof. The results follow from known results for the generalized normal distribution, but we will give separate proofs for completeness. Note that \( \alpha^{-1/n} \) is a scale parameter of the family of distributions. That is, if \( Z \) has the “standard” distribution with parameters \( n \) and 1, and with density function \( z \mapsto e^{-b_n |z|^n} \) then \( X = \alpha^{-1/n} Z \) has the distribution with parameters \( n \) and \( \alpha \) and with density function \( g \).

(a) When \( n = 1 \),

\[
g(x) = \alpha e^{-2\alpha |x|}, \quad x \in \mathbb{R}
\]

(b) When \( n = 2 \),

\[
g(x) = \sqrt{\alpha} e^{-\alpha \pi x^2}, \quad x \in \mathbb{R}
\]

(c) If \( m \in \mathbb{N}_+ \), then by a well-known integration formula,

\[
\mathbb{E}(|Z|^m) = 2 \int_0^\infty z^m e^{-b_n z^n} dz = \left( \frac{n}{2} \right)^m \frac{\Gamma(m+1)}{\Gamma(m+1) \left( \frac{1}{n} \right)}
\]

If \( m \) is odd, then \( \mathbb{E}(Z^m) = 0 \) since the distribution is symmetric about 0. If \( m \) is even then of course \( \mathbb{E}(|Z|^m) = \mathbb{E}(Z^m) \). Finally, \( E(X^m) = \alpha^{-m/2} \mathbb{E}(Z^m) \).
(d) Again, it suffices to consider the standard variable $Z$, since $\alpha^{-1/n} \to 1$ as $n \to \infty$. But $e^{-b_n|z|^n} \to 1$ as $n \to \infty$ if $|z| < \frac{1}{2}$ and $e^{-b_n|z|^n} \to 0$ as $n \to \infty$ if $|z| \geq \frac{1}{2}$.

In particular, the variance and kurtosis are given by

$$\text{var}(X) = \alpha^{-n/2} \left(\frac{n}{2}\right)^2 \frac{\Gamma(3/n)}{\Gamma(1/n)} \quad \text{kurt}(X) = \frac{\Gamma(5/n)\Gamma(1/n)}{\Gamma^2(3/n)}, \quad n \in \mathbb{N}_+$$

This class of distributions on $\mathbb{R}$ is interesting because it generalizes the Laplace, the $0$ mean normal, and because of the surprising convergence result in part (d). Graphs of $g$ for $n \in \{1, 2, 5\}$ are given in Figure 10.1. Since the distribution of $X$ is symmetric about 0, it also follows that $|X|$ has density function $2g$ on $[0, \infty)$, $\text{sgn}(X)$ is uniformly distributed on $\{-1, 1\}$, and that $|X|$ and $\text{sgn}(X)$ are independent.

![Graphs of g](image)

**Figure 10.1:** Graphs of $g$ for $n = 1, 2, 5$

We can rephrase part (a) of Corollary 10.5: If $X$ has the Laplace distribution on $\mathbb{R}$ with scale parameter $a \in (0, \infty)$, then $X$ has constant rate $\alpha = 1/(2a)$ for the graph $(\mathbb{R}, \Rightarrow)$. Similarly we can rephrase (b): If $X$ and $Y$ are independent, and each has the normal distribution with mean $0$ and standard deviation $\sigma \in (0, \infty)$, then $(X, Y)$ has constant rate $\alpha = 1/(2\pi\sigma^2)$ for the graph $(\mathbb{R}^2, \Rightarrow_2)$.

Part (d) of Corollary 10.5 suggests that the uniform distribution on $\left[ -\frac{1}{2}, \frac{1}{2} \right]$ has constant rate for the graph corresponding to $n = \infty$. That is indeed true, properly understood. Let

$$\mathbb{R}^\infty = \{ \mathbf{x} = (x_1, x_2, \ldots) : x_i \in \mathbb{R} \text{ for each } i \in \mathbb{N}_+ \}$$

the infinite power space of $\mathbb{R}$. The corresponding $\sigma$-algebra $\mathcal{B}^\infty$ and reference measure $\lambda^\infty$ are defined as in Chapter 1. Next, define

$$\|\mathbf{x}\|_\infty = \sup\{|x_i| : i \in \mathbb{N}_+\}, \quad \mathbf{x} \in \mathbb{R}^\infty$$

$$l_\infty = \{ \mathbf{x} \in \mathbb{R}^\infty : \|\mathbf{x}\|_\infty < \infty \}$$

Note that $\mathbf{x} \in l_\infty$ if and only if $|x_i|$ is bounded in $i \in \mathbb{N}_+$. So $l_\infty$ is a measurable subset of $\mathbb{R}^\infty$ and is given the induced $\sigma$-algebra and reference measure. The graph $(l_\infty, \Rightarrow_\infty)$ induced by the function $\mathbf{x} \to \|\mathbf{x}\|_\infty$ and the standard graph $([0, \infty), \leq)$ is defined in the usual way: $\mathbf{x} \Rightarrow_\infty \mathbf{y}$ if and only if $\|\mathbf{x}\|_\infty \leq \|\mathbf{y}\|_\infty$ for $\mathbf{x}, \mathbf{y} \in l_\infty$.

**Theorem 10.13.** Suppose that $\mathbf{X} = (X_1, X_2, \ldots)$ is a sequence of independent variables, each uniformly distributed on $[-\frac{1}{2}, \frac{1}{2}]$. Then $\mathbf{X}$ has constant rate $1$ for the graph $(l_\infty, \Rightarrow_\infty)$.

**Proof.** Random variable $X_i$ has density function $g$ on $\mathbb{R}$ defined by $g(x) = 1$ for $x \in [-\frac{1}{2}, \frac{1}{2}]$ and $g(x) = 0$ otherwise. Hence $\mathbf{X}$ had density function $f$ on $l_\infty$ defined by $f(\mathbf{x}) = 1$ for $\mathbf{x} \in [-\frac{1}{2}, \frac{1}{2}]^\infty$ and $f(\mathbf{x}) = 0$ otherwise. On the other hand, the right probability function $F$ of $\mathbf{X}$ for $(l_\infty, \Rightarrow_\infty)$ is

$$F(\mathbf{x}) = \mathbb{P}(\mathbf{x} \Rightarrow_\infty \mathbf{X}) = \mathbb{P}(\|\mathbf{x}\|_\infty \leq \|\mathbf{X}\|_\infty), \quad \mathbf{x} \in l_\infty$$

But $\mathbb{P}(\|\mathbf{X}\|_\infty = \frac{1}{2}) = 1$. To see this, note that if $\epsilon > 0$ then with probability $1$, $|X_i| > \frac{1}{2} - \epsilon$ for some (and in fact infinitely many) $i \in \mathbb{N}_+$. Hence, on $l_\infty$, $F(\mathbf{x}) = 1$ if $\mathbf{x} \in [-\frac{1}{2}, \frac{1}{2}]^\infty$ and $F(\mathbf{x}) = 0$ otherwise. That is, $F = f$. 


Returning to the general case of $n \in \mathbb{N}_+$ and $k \in [1, \infty]$, consider the random walk $(X_1, X_2, \ldots)$ on $(\mathbb{R}^n, \Rightarrow_k)$ associated with $X$, where $X$ has constant rate $\alpha \in (0, \infty)$ for $(\mathbb{R}^n, \Rightarrow_k)$. In addition to our usual general results, we can explicitly give the density function $f_m$ of $X_m$ for $m \in \mathbb{N}_+$:

$$f_m(x) = \alpha^m \frac{1}{(m-1)!} (b_{n,k}\|x\|_k^n)^{m-1} \exp(-\alpha b_{n,k}\|x\|_k^n), \quad x \in \mathbb{R}^n$$

Note that $f_m(x)$ is the ordinary gamma density function for $([0, \infty), \leq)$, (with order $m$ and rate $\alpha$), evaluated at $b_{n,k}\|x\|_k^n$.  

Chapter 11

Standard Discrete Spaces

Our primary goal in this chapter is to study the discrete positive semigroup \((\mathbb{N}, +)\) with its associated graph \((\mathbb{N}, \leq)\). This space is the primary model of discrete time in many applications. In addition, we will study some closely related graphs on \(\mathbb{N}\):

(a) The graph \((\mathbb{N}, <)\) which is the reflexive deletion of \((\mathbb{N}, \leq)\).
(b) The covering graph \((\mathbb{N}, \uparrow)\) of \((\mathbb{N}, \leq)\), or a bit more generally, the graph \((\mathbb{N}, \uparrow^n)\) where \(\uparrow^n\) is the composition power of \(\uparrow\) of order \(n \in \mathbb{N}\).
(c) The reflexive completion \((\mathbb{N}, \uparrow)\) of \((\mathbb{N}, \uparrow)\).

There are also corresponding spaces on \(\mathbb{N}^+\) that are isomorphic to the spaces above, and so we do not need to study these separately. In particular,

(a) The strict positive semigroup \((\mathbb{N}^+, +)\) with associated order <. The graph \((\mathbb{N}^+, <)\) is isomorphic to \((\mathbb{N}, <)\).
(b) The reflexive completion \((\mathbb{N}^+, \leq)\) of \((\mathbb{N}^+, <)\), isomorphic to \((\mathbb{N}, \leq)\).
(c) The covering graph \((\mathbb{N}^+, \uparrow)\) for \((\mathbb{N}^+, \leq)\), isomorphic to \((\mathbb{N}, \uparrow)\).
(d) The reflexive completion \((\mathbb{N}^+, \uparrow)\) of \((\mathbb{N}^+, \uparrow)\), isomorphic to \((\mathbb{N}, \uparrow)\).

As usual, counting measure \(\#\) is the reference measure. For the semigroup \((\mathbb{N}, +)\) it is the unique left invariant measure, up to multiplication by positive constants. The main results are given for the positive semigroup \((\mathbb{N}, +)\). Most of these results are well known, but perhaps not in this context. Most of the results for the other graphs are left as exercises.

11.1 Basics

As usual, we start with the left walk functions and the corresponding generating function.

**Proposition 11.1.** The left walk function \(w_n\) of order \(n \in \mathbb{N}\) for \((\mathbb{N}, \leq)\) is given by

\[
w_n(x) = \binom{x + n}{n}, \quad x \in \mathbb{N}
\]

**Proof.** For a proof by induction, note that the expression is correct when \(n = 0\). Assume that the expression is correct for a given \(n \in \mathbb{N}\). Then

\[
w_{n+1}(x) = \sum_{z=0}^{x} w_n(z) = \sum_{z=0}^{x} \binom{n + z}{x} = \binom{n + 1 + x}{x}
\]

by a binomial identity. For a combinatorial proof, to construct a walk of length \(n\) in \((\mathbb{N}, \leq)\) ending in \(x\) we need to select a multiset of size \(n\) from \(\{0, 1, \ldots, x\}\). The number of ways to do this is \(\binom{x+n}{n}\). \(\Box\)

106
Exercise 11.1. Find the left walk function $w_n$ of order $n \in \mathbb{N}$ for each of the following graphs:

(a) $(\mathbb{N}, <)$  
(b) $(\mathbb{N}, \uparrow)$  
(c) $(\mathbb{N}, \Uparrow)$  

Proposition 11.2. The left generating function $W$ for $(\mathbb{N}, \leq)$ is given by

$$W(x, t) = \frac{1}{(1 - t)^{x+1}}, \quad x \in \mathbb{N}, \; |t| < 1$$

Proof. From the left walk function and a binomial identity,

$$W(x, t) = \sum_{n=0}^{\infty} \binom{n + x}{x} t^n = \frac{1}{(1 - t)^{x+1}}, \quad |t| < 1, \; x \in \mathbb{N}$$

Exercise 11.2. Find the left generating function for each of the following graphs.

(a) $(\mathbb{N}, <)$  
(b) $(\mathbb{N}, \uparrow)$  
(c) $(\mathbb{N}, \Uparrow)$  

Proposition 11.3. The Möbius kernel $M$ of $(\mathbb{N}, \leq)$ is given as follows: For $(x, y) \in \mathbb{N}^2$, $M(x, y) = 1$ if $y = x$, $M(x, y) = -1$ if $y = x + 1$, and $M(x, y) = 0$ otherwise.

Proof. This is well known, but a proof is also easy from the definition: For $x \in \mathbb{N}, M(x, x) = 1$ and

$$M(x, y) = -\sum_{t=x}^{y-1} M(x, t), \quad x < y$$

11.2 Distributions

Suppose that $X$ is a random variable with values in $\mathbb{N}$ and with probability density function $f$. As usual, when a particular graph is under discussion, we assume that the graph supports the distribution of $X$.

Proposition 11.4. For the graph $(\mathbb{N}, \leq)$, the right probability function $F$ and the right rate function $r$ of $X$ are given by

$$F(x) = \mathbb{P}(x \leq X) = \sum_{y=x}^{\infty} f(y), \quad x \in \mathbb{N}$$

$$r(x) = \frac{f(x)}{F(x)} = \frac{f(x)}{\sum_{y=x}^{\infty} f(y)}, \quad x \in \mathbb{N}$$

Both functions uniquely determine the distribution.
Proof. The formulas for the right probability function and the right rate function follow from the definitions. The density function $f$ can be recovered from $F$ by

$$f(x) = F(x) - F(x + 1), \quad x \in \mathbb{N}$$

The density function can be recovered from $r$ recursively by $f(0) = r(0)$ and

$$f(x) = r(x) \left[ 1 - \sum_{n=0}^{x-1} f(n) \right], \quad x \in \mathbb{N}_+$$

Exercise 11.3. Find the right probability function $F$ and right rate function $r$ of $X$ in terms of the density function $f$ for each of the following graphs. Show that $F$ uniquely determines $f$ in each case.

(a) $(\mathbb{N}, <)$
(b) $(\mathbb{N}, \uparrow)$,
(c) $(\mathbb{N}, \updownarrow)$,

✓✓

The standard moment result for the graph $(\mathbb{N}, \leq)$ is given next.

Proposition 11.5. Suppose again that $X$ is a random variable with values in $\mathbb{N}$. Then

$$\sum_{x=0}^{\infty} \binom{n+x}{n} \mathbb{P}(X \geq x) = \mathbb{E} \left[ \binom{n+1+X}{n+1} \right], \quad n \in \mathbb{N}$$

When $n = 0$, this is equivalent to the well-known result

$$\mathbb{E}(X) = \sum_{x=0}^{\infty} \mathbb{P}(X > x) = \sum_{x=1}^{\infty} \mathbb{P}(X \geq x)$$

Exercise 11.4. Explicitly give the standard moment result for each of the following graphs:

(a) $(\mathbb{N}, <)$
(b) $(\mathbb{N}, \uparrow)$
(c) $(\mathbb{N}, \updownarrow)$

✓✓

Of course, we are most interested in random variables with the exponential, memoryless, and constant rate properties. We will prove the famous result for the semigroup $(\mathbb{N}, +)$ and leave the proofs for the other graphs as exercises. Suppose again that $X$ is a random variable with values in $\mathbb{N}$.

Theorem 11.1. Random variable $X$ is exponential for $(\mathbb{N}, +)$ if and only if it is memoryless for $(\mathbb{N}, +)$ if and only if it has constant rate for $(\mathbb{N}, \leq)$. The exponential distribution on $(\mathbb{N}, +)$ with constant rate $\alpha \in (0, 1)$ has density function $f$ given by

$$f(x) = \alpha(1 - \alpha)^x, \quad x \in \mathbb{N}$$

Proof. Let $F$ denote the right probability function of $X$. The memoryless property is

$$F(x + y) = F(x)F(y), \quad x, y \in \mathbb{N}$$

The only solution is $F(x) = (1 - \alpha)^x$ where $\alpha = 1 - F(1)$. The constant rate properties reduces to $f(0) = \alpha$ and

$$F(x + 1) = (1 - \alpha)F(x), \quad x \in \mathbb{N}$$

so once again, $F(x) = (1 - \alpha)^x$ for $x \in \mathbb{N}$. □
Exercise 11.5. Show that $X$ has constant rate for $(\mathbb{N}, \leq)$ if and only if $X + 1$ is exponential for $(\mathbb{N}^+, +)$ if and only if $X + 1$ is memoryless for $(\mathbb{N}^+, +)$. The distribution with constant rate $\alpha \in (0, \infty)$ for $(\mathbb{N}, \leq)$ has density function $f$ given by

$$f(x) = \frac{\alpha}{1 + \alpha} \left( \frac{1}{1 + \alpha} \right)^x, \quad x \in \mathbb{N}$$

Exercise 11.6. Show that $X$ has constant rate $\alpha \in (1, \infty)$ for $(\mathbb{N}, \uparrow)$ if and only if $X$ has density function $f$ given by

$$f(x) = \frac{\alpha - 1}{\alpha} \left( \frac{1}{\alpha} \right)^x, \quad x \in \mathbb{N}$$

and in this case, $X$ has constant rate $\alpha^n$ for $(\mathbb{N}, \uparrow^n)$ for each $n \in \mathbb{N}$.

Exercise 11.7. Show that $X$ has constant rate $\alpha \in (1/2, 1)$ for $(\mathbb{N}, \uparrow)$ if and only if $X$ has density function $f$ given by

$$f(x) = \frac{2\alpha - 1}{\alpha} \left( \frac{1 - \alpha}{\alpha} \right)^x, \quad x \in \mathbb{N}$$

Of course, each of these distributions is a geometric distribution. In a sequence of Bernoulli trials, the geometric distribution on $\mathbb{N}$ governs the number of failures before the first success. The distribution in Theorem 11.1 is geometric with success parameter $\alpha$. The distribution in Exercise 11.5 is geometric with success parameter $\alpha/(1+\alpha)$. The distribution in Exercise 11.6 is geometric with success parameter $(\alpha - 1)/\alpha$. The distribution in Exercise 11.7 is geometric on $\mathbb{N}$ with success parameter $(2\alpha - 1)/\alpha$. Here is another way of looking at the results.

Corollary 11.1. Suppose that $X$ has the geometric distribution on $\mathbb{N}$ with success parameter $p \in (0, 1)$, so that $X$ has density function $f$ given by $f(x) = p(1-p)^x$ for $x \in \mathbb{N}$. Then

(a) $X$ has constant rate $p$ for $(\mathbb{N}, \leq)$.
(b) $X$ has constant rate $p/(1-p)$ for $(\mathbb{N}, <)$.
(c) $X$ has constant rate $1/(1-p)^n$ for $(\mathbb{N}, \uparrow^n)$ for each $n \in \mathbb{N}$.
(d) $X$ has constant rate $1/(2-p)$ for $(\mathbb{N}, \uparrow)$.

Theorem 11.2. Suppose that $X$ has the exponential distribution on $(\mathbb{N}, +)$ with constant rate $p \in (0, 1)$. Then $X$ has a compound Poisson distribution. Specifically, $X$ can be decomposed as

$$X = U_1 + U_2 + \cdots + U_N$$

where $U = (U_1, U_2, \ldots)$ is a sequence of independent variables, each with the logarithmic distribution on $\mathbb{N}$:

$$P(U = n) = -\frac{(1-p)^n}{n \ln p}, \quad n \in \mathbb{N}$$

and where $N$ is independent of $U$ and has the Poisson distribution with parameter $-\ln p$.

Proof. The exponential distribution on $(\mathbb{N}, +)$ with rate $p \in (0, 1)$ is of course the geometric distribution with success parameter $p$. It’s well known that the geometric distribution has the decomposition stated in the theorem. Here is a direct proof: The common probability generating function $P$ of $U_i$ for $i \in \mathbb{N}$ is given by

$$P(t) = \frac{\ln[1 - (1-p)t]}{\ln p}, \quad |t| < \frac{1}{1-p}$$

The probability generating function $Q$ of the $N$ is given by

$$Q(t) = \exp[-\ln(p)(t-1)], \quad t \in \mathbb{R}$$
Hence the probability generating function of $U_1 + U_2 + \cdots + U_N$ is $Q \circ P$, given by

$$Q[P(t)] = \frac{1}{1 - (1 - p)t}, \quad |t| < \frac{1}{1 - p}$$

which we recognize as the probability generating function of the geometric distribution with success parameter $p$. \hfill \Box

Theorem 11.2 shows how an exponential variable for $(\mathbb{N}, +)$ can be written as a Poisson random sum of independent, identically distributed variables. The following result shows how the exponential variable can be written in terms of an infinite sum of independent Poisson variables.

**Theorem 11.3.** Suppose that $X = (X_1, X_2, \ldots)$ is a sequence of independent variables and that $X_n$ has the Poisson distribution on $\mathbb{N}$ with parameter $\lambda_n = (1 - p)^n/n$ for $n \in \mathbb{N}_+$ where $p \in (0, 1)$. Then $Y = \sum_{n=1}^{\infty} nX_n$ has the exponential distribution on $(\mathbb{N}, +)$ with rate $p$.

**Proof.** The probability generating function of $X_n$ is

$$\mathbb{E}(t^{X_n}) = \exp[\lambda_n(t - 1)] = \exp \left[ \frac{(1 - p)^n}{n} (t - 1) \right]$$

Hence the probability generating function of $Y$ is

$$\mathbb{E}(t^Y) = \prod_{n=1}^{\infty} \mathbb{E}(t^{nX_n}) = \prod_{n=1}^{\infty} \exp \left[ \frac{(1 - p)^n}{n} (t^n - 1) \right]$$

$$= \exp \left[ \sum_{n=1}^{\infty} \frac{(1 - p)^n}{n} (t^n - 1) \right] = \exp[-\ln(1 - (1 - p)t) + \ln(p)]$$

$$= \frac{p}{1 - (1 - p)t}, \quad |t| < \frac{1}{1 - p}$$

But this is the probability generating function of the geometric distribution with rate parameter $p$. \hfill \Box

**Proposition 11.6.** If $X$ has constant rate $\alpha \in (0, 1)$ for the graph $(\mathbb{N}, \leq)$ then $X$ maximizes entropy over all random variables $Y$ on $\mathbb{N}$ with $\mathbb{E}(Y) = (1 - \alpha)/\alpha$.

**Proof.** From the general theory, the distribution of a random variable $X$ with constant rate $\alpha$ for a graph $(S, \rightarrow)$ maximizes entropy over all random variables $Y$ with values in $S$ and $\mathbb{E}(\ln[F(Y)]) = \mathbb{E}(\ln[F(X)])$. For the graph considered here, this condition reduces to $\mathbb{E}(Y) = \mathbb{E}(X) = (1 - \alpha)/\alpha$. \hfill \Box

### 11.3 Random Walks

Our main interest in this section is the random walk on $(\mathbb{N}, +)$ (equivalently the random walk on $(\mathbb{N}, \leq)$) associated with the exponential distribution on $(\mathbb{N}, +)$ (equivalently the constant rate distribution on $(\mathbb{N}, \leq)$). Recall that the random walk $X = (X_1, X_2, \ldots)$ on $(\mathbb{N}, +)$ can be constructed from an IID sequence $U = (U_1, U_2, \ldots)$ with the underlying distribution by means of partial sums. So in this case, the term random walk has its elementary meaning. The random walk $X$ on $(\mathbb{N}, \leq)$ can be constructed from $U$ by means of record variables. Recall also that the random walk $X$ corresponding to a constant rate distribution is the “most random” way to put points in the graph in the sense that given $X_{n+1} = x$, the sequence $(X_1, X_2, \ldots, X_n)$ is uniformly distributed on the set of walks of length $n$ in the graph that terminate in $x$.

**Theorem 11.4.** Suppose that $X = (X_1, X_2, \ldots)$ is the random walk on $(\mathbb{N}, +)$ (equivalently $(\mathbb{N}, \leq)$) corresponding to the exponential distribution with constant rate $\alpha \in (0, 1)$.

(a) The transition density $P^n$ of order $n \in \mathbb{N}$ for $X$ is given by

$$P^n(x, y) = \binom{y - x + n - 1}{n - 1} \alpha^n(1 - \alpha)^{y-x}, \quad x, y \in \mathbb{N}, x \leq y$$
(b) For \( n \in \mathbb{N}_+ \), \((X_1, X_2, \ldots, X_n)\) has density function \( g_n \) defined by
\[
g_n(x_1, x_2, \ldots, x_n) = \alpha^n (1 - \alpha)^{x_n}, \quad (x_1, x_2, \ldots, x_n) \in \mathbb{N}^n, \ x_1 \leq x_2 \leq \cdots \leq x_n
\]

(c) For \( n \in \mathbb{N}_+ \), \( X_n \) has density function \( f_n \) defined by
\[
f_n(x) = \alpha^n \binom{n + x - 1}{x} (1 - \alpha)^x, \quad x \in \mathbb{N}
\]

(d) Given \( X_{n+1} = x \in \mathbb{N} \), \((X_1, X_2, \ldots, X_n)\) is uniformly distributed on the \( \binom{n + x}{n} \) points in the set
\[
\{(x_1, x_2, \ldots, x_n) \in \mathbb{N}^n : x_1 \leq x_2 \leq \cdots \leq x_n \leq x\}
\]

Of course we recognize the distribution of \( X_n \) as the negative binomial distribution on \( \mathbb{N} \) with parameters \( n \) and \( \alpha \). This distribution governs the number of failures before the \( n \)th success in a sequence of Bernoulli trials with success parameter \( \alpha \).

**Exercise 11.8.** Give the results analogous to Theorem 11.4 for the graph \((\mathbb{N}, <)\).

**Exercise 11.9.** Give the results analogous to Theorem 11.4 for the graph \((\mathbb{N}, \uparrow)\).

**Exercise 11.10.** Give the results analogous to Theorem 11.4 for the graph \((\mathbb{N}, \Downarrow)\).

Next we consider the point process associated with a random walk \( X = (X_1, X_2, \ldots) \) on \((\mathbb{N}, \leq)\). Specifically, let \( N_x = \#\{n \in \mathbb{N}_+ : X_n \leq x\} \) for \( x \in \mathbb{N} \). We are interested in the function that plays the role that the renewal function plays in ordinary renewal theory, that is, \( x \mapsto E(N_x) \).

**Theorem 11.5.** Suppose again that \( X = (X_1, X_2, \ldots) \) is the random walk on \((\mathbb{N}, \leq)\) (equivalently \((\mathbb{N}, \leq)\)) corresponding to the exponential distribution with constant rate \( \alpha \in (0, 1) \). Then
\[
E(N_x) = \frac{\alpha}{1 - \alpha} (x + 1), \quad x \in \mathbb{N}
\]

**Proof.** Recall that if \( X \) is the random walk on \((\mathbb{N}, \leq)\) corresponding to the distribution of \( X \), where \( X \) has constant rate \( \alpha \in (0, 1) \) then
\[
E(N_x) = E[W(X, \alpha); X \leq x]
\]
where \( W \) is the left generating function of \((\mathbb{N}, \leq)\). Hence
\[
E(N_x) = E \left[ \frac{1}{(1 - \alpha)^x}; X \leq x \right] = \sum_{t=0}^{x} \frac{1}{(1 - \alpha)^{t+1}} \alpha^{t+1} = \frac{\alpha}{1 - \alpha} (x + 1), \quad x \in \mathbb{N}
\]

**Exercise 11.11.** Give results analogous to Theorem 11.5 for each of the following graphs:

(a) \((\mathbb{N}, <)\)

(b) \((\mathbb{N}, \uparrow)\)

(c) \((\mathbb{N}, \Downarrow)\)

Suppose again that \( X = (X_1, X_2, \ldots) \) is a random walk on one of the graphs considered in this section. Suppose that \( N \) is independent of \( X \) and has the geometric distribution on \( \mathbb{N}_+ \) with success probability \( q \in (0, 1) \), so that \( \mathbb{P}(N = n) = q(1 - q)^{n-1} \) for \( n \in \mathbb{N}_+ \). Recall that this distribution is exponential for the semigroup \((\mathbb{N}_+, +)\), but with rate \( q/(1 - q) \). We accept or reject each point in \( X \), independently, with probabilities \( q \) and \( 1 - q \), respectively. So \( Y_X \) is the first accepted point.

**Theorem 11.6.** Suppose that \( X \) is the random walk on \((\mathbb{N}, \leq)\) corresponding to the distribution with constant rate \( \alpha \in (0, 1) \). Then \( X_N \) has constant rate \( \alpha q/(1 - \alpha + \alpha q) \) for \((\mathbb{N}, \leq)\), with density function \( g \) given by
\[
g(x) = \frac{\alpha q}{1 - \alpha + \alpha q} \left( \frac{1 - \alpha}{1 - \alpha + \alpha q} \right)^x, \quad x \in \mathbb{N}
\]
Proof. Recall that
\[ g(x) = qaW[x, (1 - q)\alpha]F(x), \quad x \in S \]
where \( W \) is the left generating function and \( F \) is the right probability function of the distribution for the graph. For the graph \((\mathbb{N}, \leq)\), the left generating function \( W \) is given by \( W(x, t) = 1/(1 - t)^{x+1} \) for \( x \in \mathbb{N} \) and \( t \in (-1, 1) \) and the right probability function \( F \) is given by \( F(x) = (1 - \alpha)^x \) for \( x \in \mathbb{N} \). Substituting gives the result. \( \square \)

Exercise 11.12. Find results analogous to Theorem 11.6 for each of the following graphs:

(a) \((\mathbb{N}, <)\)
(b) \((\mathbb{N}, \uparrow)\)
(c) \((\mathbb{N}, \uparrow\downarrow)\)

\( \checkmark \) \( \checkmark \)

Our last results in this section deal with compound Poisson distributions.

Corollary 11.2. Suppose that \( X = (X_1, X_2, \ldots) \) is the random walk on \((\mathbb{N}, +)\) corresponding to the exponential distribution with constant rate \( p \in (0, 1) \). For \( n \in \mathbb{N}_+ \), \( X_n \) has a compound Poisson distribution. Specifically,
\[ X_n = W_1 + W_2 + \cdots + W_N \]
where \((W_1, W_2, \ldots)\) is a sequence of independent variables, each with the logarithmic distribution on \( \mathbb{N} \) with parameter \( p \), and where \( N \) is independent of \( W \) and has the Poisson distribution with parameter \(-n \ln p\).

Proof. This follows immediately from Theorem 11.2 and Proposition 5.4. \( \square \)

11.4 Quotient spaces

Consider again the positive semigroup \((\mathbb{N}, +)\). For \( t \in \mathbb{N}_+ \), the sub-semigroup generated by \( t \) is
\[ t\mathbb{N} = \{nt : n \in \mathbb{N}\} \]
and the corresponding quotient space is \( \{0, \ldots, t-1\} \). The general assumptions in Example 9.1 hold, so that \( x \in \mathbb{N} \) can be written uniquely as
\[ x = tn_t(x) + z_t(x) \]
where \( n_t(x) = |x/t| \in \mathbb{N} \) and \( z_t(x) = x - y_t(x) = x \mod t \in \{0, 1, \ldots, t-1\} \). The following theorem follows from the general results in Chapter 9.

Theorem 11.7. Suppose that \( X \) has the exponential distribution on \((\mathbb{N}, +)\) with rate \( p \in (0, 1) \). Then

(a) \( N_t \) and \( Z_t \) are independent.

(b) \( N_t \) has the exponential distribution on \((\mathbb{N}, +)\) with rate \( 1 - (1 - p)^t \), and this is also the conditional distribution of \( X/t \) given \( X \in t\mathbb{N} \).

(c) \( Z_t \) has probability density function \( k \mapsto p(1 - p)^k/[1 - (1 - p)^t] \) on \( \{0, \ldots, t-1\} \), and this is the conditional distribution of \( X \) given \( X \in \{0, 1, \ldots, t-1\} \).

As before, we are interested in a converse that is stronger than the general converse in Theorem 9.4. Thus, suppose that \( X \) is a random variable taking values in \( \mathbb{N} \), with density function \( f \) and right probability function \( F \) for \((\mathbb{N}, \leq)\). Then the independence of \( Z_t \) and \( \{N_t = 0\} \) is equivalent to
\[ f(k) = [1 - F(t)] \sum_{n=0}^{\infty} f(nt + k) \quad (11.1) \]
for all \( t \in \mathbb{N}_+ \) and all \( k \in \{0, \ldots, t-1\} \). However, as in the continuous case, it is easy to see that if \( X \) has a geometric distribution, then (11.1) holds for all \( t \in \mathbb{N}_+ \) and \( k \in \mathbb{N} \), not just \( k \in \{0, 1, \ldots, t-1\} \).
Theorem 11.8. Suppose that (11.1) holds for \( k = 0 \) and for \( k = t \), for all \( t \in \mathbb{N}_+ \). Then \( X \) has an exponential distribution on \( (\mathbb{N}, +) \).

Proof. The hypotheses are

\[
\begin{align*}
  f(0) &= [1 - F(t)] \sum_{n=0}^{\infty} f(nt), \quad t \in \mathbb{N}_+ \\
  f(t) &= [1 - F(t)] \sum_{n=0}^{\infty} f((n+1)t), \quad t \in \mathbb{N}_+
\end{align*}
\]

(11.2) and (11.3)

From these equations, it follows that

\[
f(t) = f(0) F(t), \quad t \in \mathbb{N}_+
\]

and hence \( X \) has constant rate \( p = f(0) \) for \( (\mathbb{N}, \leq) \). Equivalently \( X \) has the exponential distribution for \( (\mathbb{N}, +) \) with rate \( p \) which of course is the geometric distribution with success parameter \( p \).

However, there are non-geometrically distributed variables for which \( N_t \) and \( Z_t \) are independent. The following result is easy to verify.

Proposition 11.7. Suppose that the support of \( X \) is \( \{0, 1\} \) or \( \{0, 2\} \) or \( \{c\} \) for some \( c \in \mathbb{N} \). Then \( N_t \) and \( Z_t \) are independent for all \( t \in \mathbb{N}_+ \).

The following theorem gives a partial converse:

Proposition 11.8. Suppose that \( X \) takes values in a proper subset of \( \mathbb{N} \) and that \( Z_t \) and \( \{N_t = 0\} \) are independent for all \( t \in \mathbb{N}_+ \). Then the support of \( X \) is one of the sets in Proposition 11.7.

Proof. Let \( f \) denote the density function and let \( F \) denote the right probability function of \( X \) for \( (\mathbb{N}, \leq) \).

First, we will use induction on \( a \in \mathbb{N} \) to show that if \( X \) takes values in \( \{0, 1, \ldots, a\} \), then the support of \( X \) is one of the sets in Proposition 11.7. If \( a = 0 \) or \( a = 1 \), the result is trivially true. Suppose that the statement is true for a given \( a \in \{2, 3, \ldots\} \), and suppose that \( X \) takes values in \( \{0, 1, \ldots, a+1\} \). With \( t = a+1 \), (11.1) becomes

\[
f(k) = \sum_{j=0}^{a} f(j) \sum_{n=0}^{\infty} f[n(a+1) + k], \quad k \in \{0, \ldots, a\}
\]

(11.4)

But \( \sum_{j=0}^{a} f(j) = 1 - f(a+1) \). Hence (11.4) gives

\[
\begin{align*}
  f(0) &= [1 - f(a+1)][f(0) + f(a+1)], \quad (k = 0) \\
  f(k) &= [1 - f(a+1)]f(k), \quad k \in \{1, \ldots, a\}
\end{align*}
\]

(11.5) and (11.6)

Suppose that \( f(k) > 0 \) for some \( k \in \{1, \ldots, a\} \). Then from (11.6), \( f(a+1) = 0 \). Hence \( X \) takes values in \( \{0, \ldots, a\} \), so by the induction hypothesis, the support set of \( X \) is \( \{0, 1\} \), \( \{0, 2\} \), or \( \{c\} \) for some \( c \). Suppose that \( f(k) = 0 \) for all \( k \in \{1, \ldots, a\} \). Thus, \( X \) takes values in \( \{0, a+1\} \). But then (11.4) with \( t = a \) and \( k = 0 \) gives \( f(0) = f(0)f(0) \). If \( f(0) = 0 \) then the support of \( X \) is \( \{a+1\} \). If \( f(0) = 1 \) then the support of \( X \) is \( \{0\} \).

To complete the proof, suppose that \( X \) takes values in a proper subset of \( \mathbb{N} \), so that \( f(k) = 0 \) for some \( k \in \mathbb{N} \). Then \( 1 - F(t) = P(X < t) > 0 \) for \( t \) sufficiently large and hence by (11.1), \( f(t+k) = 0 \) for \( t \) sufficiently large. Thus \( X \) takes values in \( \{0, 1, \ldots, a\} \) for some \( a \), and hence the support of \( X \) is one of the sets in Proposition 11.7.

Problem 11.1. If \( X \) has support \( \mathbb{N} \) and if \( Z_t \) and \( \{N_t = 0\} \) are independent for each \( t \in \mathbb{N}_+ \), (equivalently, (11.1) holds for \( t \in \mathbb{N}_+ \) and \( k \in \{0, 1, \ldots, t-1\} \)) does \( X \) have a geometric distribution?
Chapter 12

Rooted Trees

12.1 Basics

Suppose that \((S, \uparrow)\) is a (discrete) rooted, directed tree with root \(e\), and with the property that there is a finite path from \(e\) to \(x\) for each \(x \in S\). Our primary interest is in the partially ordered graph \((S, \preceq)\) that has \((S, \uparrow)\) as its covering graph, so that \(x \preceq y\) if and only if \(x = y\) or there is a path in \((S, \uparrow)\) from \(x\) to \(y\). But the analysis of \((S, \preceq)\) depends very much on the \((S, \uparrow)\), and more generally on \((S, \uparrow^n)\) where \(\uparrow^n\) is the \(n\)-fold composition power of \(\uparrow\) for \(n \in \mathbb{N}\). Other graphs of secondary interest are

(a) \((S, \uparrow)\), the reflexive completion \((S, \uparrow)\), so that \(x \uparrow y\) if and only if \(x = y\) or \(x \uparrow y\) for \((x, y) \in S^2\). That is, \((S, \uparrow)\) is the graph obtained by adding a loop in the tree \((S, \uparrow)\) at each vertex \(x \in S\).

(b) \((S, \prec)\), the strict partially ordered graph corresponding to \((S, \preceq)\), so that \(\preceq\) is the reflexive completion of \(\prec\). That is, \(x \prec y\) if and only if there is a path in \((S, \uparrow)\) from \(x\) to \(y\).

The standard discrete graphs studied in Chapter 11 are special cases of the graphs considered here, where each vertex has one child. The base set \(S\) is countably infinite, since the tree has no leaves, but by assumption, \((S, \preceq)\) is locally finite, so that \([e, x]\) is finite for each \(x \in S\). But of course it’s possible for a vertex to have infinitely many children. The graph \((S, \preceq)\) is uniform since there is a unique path from \(x\) to \(y\) whenever \(x \prec y\). Recall that \(d(x, y)\) denotes the distance from \(x\) to \(y\) when \(x \preceq y\) and \(d(x, y) = \infty\) otherwise. Thus \(d(x, x) = 0\) for \(x \in S\) and \(d(x, y)\) is the length of the (unique) path in \((S, \uparrow)\) from \(x\) to \(y\) when \(x \prec y\). In particular, \(d(x) := d(e, x) \in \mathbb{N}\) for \(x \in S\), and is the distance from \(e\) to \(x\). Then also \(d(x, y) = d(y) - d(x)\) if \(x \preceq y\). Note also that for \(d(x, y) = n\) if and only if \(x \uparrow^n y\) for \((x, y) \in S^2\) and \(n \in \mathbb{N}\). Thus the collection

\[\{\{y \in S : x \uparrow^n y\} : n \in \mathbb{N}\} = \{\{y \in S : d(x, y) = n\} : n \in \mathbb{N}\}\]

partitions \(\{y \in S : x \preceq y\} = \{y \in S : d(x, y) < \infty\}\). In particular, when \(x = e\), the collection partitions \(S\).

As usual, we start with some basic graph functions.

**Proposition 12.1.** The left walk function \(\gamma_n\) of order \(n \in \mathbb{N}\) for \((S, \preceq)\) is given by

\[
\gamma_n(x) = \binom{n + d(x)}{n} = \binom{n + d(x)}{d(x)}, \quad x \in S
\]

**Proof.** Recall that by definition, \(\gamma_0(x) = 1\) for \(x \in S\). To construct a walk of length \(n\) ending in \(x\) in the graph \((S, \preceq)\), we need to select a multiset of size \(n\) from the \(d(x) + 1\) vertices on the unique path of length \(d(x)\) in \((S, \uparrow)\) form \(e\) to \(x\). The number of ways to do this is \(\binom{n + d(x) + 1 - 1}{d(x)}\). \(\square\)

**Exercise 12.1.** Find the left walk function \(\gamma_n\) of order \(n \in \mathbb{N}\) for each of the following graphs:

(a) \((S, \uparrow)\),

(b) \((S, \uparrow)\),

(c) \((S, \prec)\),
Proposition 12.2. The left generating function $\Gamma$ for $(S, \preceq)$ is given by

$$\Gamma(x, t) = \frac{1}{(1 - t)^{d(x) + 1}}, \quad x \in S, \quad |t| < 1$$

Proof. From the definition,

$$\Gamma(x, t) = \sum_{n=0}^{\infty} \binom{n + d(x)}{d(x)} t^n = \frac{1}{(1 - t)^{d(x) + 1}}, \quad |t| < 1, \quad x \in S$$

Exercise 12.2. Find the left generating function for each of the following graphs:

(a) $(S, \uparrow)$

(b) $(S, \uparrow\downarrow)$

(c) $(S, \prec)$

Proposition 12.3. The Möbius kernel $M$ of $(S, \preceq)$ is given as follows: For $(x, y) \in \mathbb{N}^2$, $M(x, y) = 1$ if $y = x$, $M(x, y) = -1$ if $x \uparrow y$, and $M(x, y) = 0$ otherwise.

Proof. This is well known, but a proof is also easy from the definition:

$$M(x, y) = -\sum_{t \in [x, y)} M(x, t), \quad x \preceq y$$

12.2 Probability

If $X$ is a random variable on $S$ with density function $f$, then by definition, the right probability function of $X$ for $(S, \preceq)$ is given by

$$F(x) = \mathbb{P}(x \preceq X) = \sum_{y \preceq y} f(y), \quad x \in S$$

The following proposition characterizes right probability functions for $(S, \preceq)$ and shows how to recover the density function.

Theorem 12.1. A function $F : S \to [0, 1]$ is a right probability function for $(S, \preceq)$ if and only if

(a) $F(e) = 1$

(b) $\sum_{x \uparrow y} F(y) \leq F(x)$ for $x \in S$.

(c) $\sum_{e \uparrow n x} F(x) \to 0$ as $n \to \infty$.

In this case, the unique density function is given by

$$f(x) = F(x) - \sum_{x \uparrow y} F(y), \quad x \in S$$

Proof. This follows from the Möbius inversion formula 2.1, using the Möbius kernel in Proposition 12.3, but we will give a direct proof. Suppose first that $F$ is the right probability function on $(S, \preceq)$ for random variable $X$ with density function $f$. Then

(a) $F(e) = \mathbb{P}(e \preceq X) = 1$
(b) \( \sum_{x \in y} F(y) = P(x < X) \leq P(x \leq X) = F(x) \) for \( x \in S \).

c) \( \sum_{e \uparrow} F(x) = P[d(X) \geq n] \rightarrow 0 \) as \( n \rightarrow \infty \).

Conversely, suppose that \( F \) satisfies the conditions in the theorem, and let \( f \) be defined by (12.1). By (b), \( f(x) \geq 0 \) for \( x \in S \). Next, for \( x \in S \),

\[
\sum_{x \leq y} f(y) = \sum_{n=0}^{\infty} \sum_{x \mid n} f(y) = \sum_{n=0}^{\infty} \sum_{x \mid n} F(y) - \sum_{y \mid z} F(z) = \lim_{m \rightarrow \infty} \sum_{n=0}^{\infty} \sum_{x \mid n} F(y) - \sum_{y \mid z} F(z)
\]

\[
= \lim_{m \rightarrow \infty} \left[ \sum_{n=0}^{m} \sum_{x \mid n} F(y) - \sum_{n=0}^{m} \sum_{x \mid n} F(z) \right] = \lim_{m \rightarrow \infty} \left[ \sum_{n=0}^{m} \sum_{x \mid n} F(y) - \sum_{n=0}^{m} \sum_{x \mid n+1} F(z) \right]
\]

Letting \( x = e \) shows that \( \sum_{x \in S} f(x) = 1 \) so that \( f \) is a probability density function. For general \( x \in S \), it then follows that \( F \) is the right probability function of \( f \).

**Proposition 12.4.** The right probability function \( F \) of a distribution for the graph \((S, \uparrow)\) uniquely determines the distribution.

**Proof.** We use Theorem 4.1, so let \( L \) denote the adjacency kernel of the graph. Suppose that \( u : S \rightarrow \mathbb{R} \) satisfies \( \sum_{x \in S} |u(x)| < \infty \), \( \sum_{x \in S} u(x) = 0 \) and \( Lu(x) = 0 \) for \( x \in S \). The last condition means that \( u(x) + \sum_{x \uparrow} u(y) = 0 \) for \( x \in S \) or equivalently \( \sum_{x \uparrow} u(y) = -u(x) \) for \( x \in S \). Recursively it follows that for \( n \in \mathbb{N} \),

\[
\sum_{x \uparrow} u(y) = (-1)^n u(x), \quad x \in S
\]

Hence

\[
\sum_{x \leq y} u(y) = \sum_{n=0}^{\infty} \sum_{x \mid n} u(y) = \sum_{n=0}^{\infty} (-1)^n u(x), \quad x \in S
\]

But the sum on the left must converge absolutely for \( x \in S \) so it follows that \( u(x) = 0 \) for \( x \in S \).

In general, there is no simple formula for recovering the density \( f \) from the right probability function \( F \) for the graph \((S, \uparrow)\). For the other two graphs under consideration, the right probability function does not uniquely determine the distribution.

**Example 12.1.** Let \( S = \{e\} \cup \{x_n : n \in \mathbb{N}_+\} \cup \{y_n : n \in \mathbb{N}_+\} \) and define the relation \( \uparrow \) on \( S \) by \( e \uparrow x_1 \), \( e \uparrow y_1 \), and for \( n \in \mathbb{N}_+ \), \( x_n \uparrow x_{n+1} \) and \( y_n \uparrow y_{n+1} \). Then \((S, \uparrow)\) is a rooted tree of the type we are studying—essentially two copies of \((\mathbb{N}, \uparrow)\) glued together at 0. Now let \( f \) be a probability density function on \( S \) with \( f(x_1) \neq f(y_1) \). Define \( g \) by \( g(x_1) = f(y_1), g(y_1) = f(x_1), \) and \( g(x) = f(x) \) for other \( x \in S \). Then \( g \) is also a density function and \( g \neq f \). The right probability functions of \( f \) and \( g \) for the graph \((S, \uparrow)\) and for the graph \((S, \lessdot)\) are the same.

**Proposition 12.5.** Suppose that random variable \( X \) takes values in \( S \). Then

\[
\sum_{x \in S} \binom{n + d(x)}{n} P(X \geq x) = \mathbb{E} \left[ \binom{n + 1 + d(X)}{n + 1} \right], \quad n \in \mathbb{N}
\]

**Proof.** From the standard moment result recall that

\[
\sum_{x \in S} \gamma_n(x) F(x) = \mathbb{E}[\gamma_{n+1}(X)], \quad n \in \mathbb{N}
\]

where \( \gamma_n \) is the left walk function of order \( n \in \mathbb{N} \) for \((S, \lessdot)\) and \( F \) is the right probability function of \( X \) for \((S, \lessdot)\). So the result follows from Proposition 12.1.
Example 12.2. Give the standard moment result for each of the following graphs:

(a) $(S, \uparrow)$
(b) $(S, \Uparrow)$
(c) $(S, \preceq)$

We will identify a class of probability distributions that have constant rate for all of the graphs under discussion, and we will identify the rate constants. A tree with leaves cannot have a constant rate distribution, so we will assume for the rest of this section that $(S, \preceq)$ has no leaves, and hence that $S$ is infinite.

Lemma 12.1. Fix $\alpha \in (0, 1)$. The following recursive scheme defines a probability density functions $f$ on $S$:

1. $f(e) = \alpha$.
2. If $f(x)$ is defined for $x \in S$, then define $f(y)$ for $x \uparrow y$ arbitrarily, subject to the restrictions

   a. $f(y) > 0$
   b. $\sum_{x \uparrow y} f(y) = (1 - \alpha)f(x)$

Proof. Since the tree has no leaves, $\{y \in S : x \uparrow y\}$ is nonempty for each $x \in S$, and there is always some way to define $f(y)$ for $x \uparrow y$ so that (a) and (b) are satisfied. In fact, if $\#\{y \in S : x \uparrow y\} \geq 2$ then this can be done in infinitely many ways. We show by induction that

$$\sum_{e \uparrow^n x} f(x) = \alpha(1 - \alpha)^n, \quad n \in \mathbb{N}$$

First, this holds for $n = 0$ by definition. If it holds for a given $n \in \mathbb{N}$, then by construction

$$\sum_{e \uparrow^{n+1} x} f(x) = \sum_{e \uparrow^n x} \sum_{x \uparrow y} f(y) = \sum_{e \uparrow^n x} (1 - \alpha)f(x) = (1 - \alpha)\alpha(1 - \alpha)^n = \alpha(1 - \alpha)^{n+1}$$

So then

$$\sum_{x \in S} f(x) = \sum_{n=0}^{\infty} \sum_{e \uparrow^n x} f(x) = \sum_{n=0}^{\infty} \alpha(1 - \alpha)^n = 1$$

Theorem 12.2. Suppose that random variable $X$ with values in $S$ has density function $f$ with parameter $\alpha \in (0, 1)$ constructed in Lemma 12.1. Then

(a) $X$ has constant rate $\alpha$ for $(S, \preceq)$.
(b) $X$ has constant rate $1/(1 - \alpha)$ for $(S, \uparrow)$.
(c) $X$ has constant rate $1/(2 - \alpha)$ for $(S, \Uparrow)$.
(d) $X$ has constant rate $\alpha/(1 - \alpha)$ for $(S, \preceq)$.

Proof. The proof is straightforward.

(a) For $(S, \preceq)$, the right probability function $F$ is given by

$$F(x) = \sum_{x \preceq y} f(y) = \sum_{n=0}^{\infty} \sum_{x \uparrow^n y} f(y)$$

But by the same proof as in Lemma 12.1,

$$\sum_{x \uparrow^n y} f(y) = f(x)(1 - \alpha)^n$$

and hence $F(x) = \sum_{n=0}^{\infty} f(x)(1 - \alpha)^n = f(x)/\alpha$ for $x \in S$. So $f = \alpha F$. 

(b) For \((S, \uparrow)\), the right probability function \(F\) is given by 
\[
F(x) = \sum_{x \uparrow y} f(y) \quad \text{for } x \in S.
\]
But by construction, 
\[
F(x) = (1 - \alpha)f(x) \quad \text{for } x \in S, \quad \text{so } f = \frac{1 - \alpha}{1 - \alpha} F.
\]
(c) For \((S, \uparrow)\), it follows from (b) and basic results on reflexive completion that the distribution has constant rate 
\[
\frac{1/(1 - \alpha)}{1 + 1/(1 - \alpha)} = 1/(2 - \alpha)
\]
(d) By (a) and basic results on reflexive completion, the distribution has constant rate \(\alpha/(1 - \alpha)\) for \((S, \prec)\).

**Corollary 12.1.** Suppose again that \(X\) has density function \(f\) with parameter \(\alpha \in (0, 1)\) constructed in Lemma 12.1, and let \(N = d(X)\). Then \(N\) has the geometric distribution on \(\mathbb{N}\) with success parameter \(\alpha\). Equivalently, \(N\) has constant rate \(\alpha\) for the graph \((\mathbb{N}, \leq)\).

**Corollary 12.2.** Suppose again that \(X\) has density function \(f\) with parameter \(\alpha \in (0, 1)\) constructed in Lemma 12.1. Then \(X\) has constant rate \(1/(1 - \alpha)^n\) for \((S, \uparrow^n)\) for each \(n \in \mathbb{N}\).

**Corollary 12.3.** Suppose next that \(X = (X_1, X_2, \ldots)\) is the random walk on \((S, \preceq)\) associated with \(X\), where \(X\) has density function \(f\) with parameter \(\alpha \in (0, 1)\) constructed in Lemma 12.1. Then \(X_n\) has density function \(f_n\) for \(n \in \mathbb{N}\) given by
\[
f_n(x) = \alpha^{n-1} \left(\frac{n - 1 + d(x)}{d(x)}\right) f(x), \quad x \in S
\]

**Exercise 12.3.** Suppose that \(X = (X_1, X_2, \ldots)\) is the random walk on each of the graphs below associated with \(X\). Find the density function \(f_n\) of \(X_n\) for \(n \in \mathbb{N}\).

(a) \((S, \uparrow)\)
(b) \((S, \uparrow)\)
(c) \((S, \prec)\)
Chapter 13

The Free Semigroup

13.1 Basics

Definition 13.1. Let \( I \) be a countable alphabet of letters, and let \( S \) denote the set of all finite length words using letters from \( I \). The empty word, with no letters, is denoted \( e \). Let \( \cdot \) denote the concatenation operation on elements of \( S \): if \( x = x_1 \cdots x_m \in S \) and \( y = y_1 \cdots y_n \in S \), then

\[
x y = x_1 \cdots x_m y_1 \cdots y_n
\]

The space \((S, \cdot)\) is a discrete, positive semigroup, known as the free semigroup generated by \( I \).

The set of words \( S \) is sometimes denoted \( I^* \), and the elements of \( I \) are the irreducible elements of \((S, \cdot)\). The adjective free is used because there are no “algebra rules” imposed on \((S, \cdot)\).

Definition 13.2. For \( x \in S \) and \( i \in I \), let \( d_i(x) \) denote the number of times that letter \( i \) occurs in \( x \), and let \( d(x) = \sum_{i \in I} d_i(x) \) denote the length of \( x \).

Note that \( d_i(x) = 0 \) for all but finitely many \( i \in I \). For the partial order graph \((S, \preceq)\) associated with \((S, \cdot)\) note that \( x \preceq y \) if and only if \( y = xt \) for some \( t \in S \), so that \( x \) is a prefix of \( y \) and then \( x^{-1} y \) is the corresponding suffix. For the cover graph \((S, \uparrow)\) of \((S, \preceq)\) note that \( x \uparrow y \) if and only if \( y = xi \) for some \( i \in I \). More generally for \( n \in \mathbb{N} \), \( x \uparrow^n y \) if and only if \( y = xt \) for some \( t \in S \) with \( d(t) = n \). Hence \((S, \uparrow)\) is a regular tree rooted at \( e \) in which each element \( x \) has \#(\( I \)) children. In particular, if \#(\( I \)) = \( k \), then \((S, \uparrow)\) is a regular \( k \)-tree. Specializing further, when \( k = 2 \) and \( I = \{0, 1\} \), the elements of \( S \) are finite bit strings and the corresponding cover graph \((S, \uparrow)\) is the binary tree. The results in Chapter 12 on rooted trees applies to the partial order graph \((S, \preceq)\), the strict partial order graph \((S, \prec)\), the covering graph \((S, \uparrow)\), and its reflexive completion \((S, \uparrow)\). In particular, \( d(x) \) is the distance from \( e \) to \( x \) in the cover graph \((S, \uparrow)\) for \( x \in S \). More generally, if \( x \preceq y \) then \( d(x,y) = d(y) - d(x) \) is the distance from \( x \) to \( y \) in \((S, \uparrow)\).

Definition 13.3. Let

\[
\mathbb{N}_I = \{(n_i : i \in I) : n_i \in \mathbb{N} \text{ for } i \in I \text{ and } n_i = 0 \text{ for all but finitely many } i \in I\}
\]

We define the multinomial coefficients on \( \mathbb{N}_I \) by

\[
C(n) = \#\{x \in S : d_i(x) = n_i \text{ for } i \in I\} = \frac{(\sum_{i \in I} n_i)!}{\prod_{i \in I} n_i!}, \quad n = (n_i : i \in I) \in \mathbb{N}_I
\]

The free semigroup has the property that \([e, x] \) is finite and totally ordered for each \( x \). Moreover, it is the only discrete positive semigroup with this property.

Proposition 13.1. Suppose that \((S, \cdot)\) is a discrete positive semigroup with associated partial order \( \preceq \) and with the property that \([e, x] \) is finite and totally ordered for each \( x \in S \). Then \((S, \cdot)\) is isomorphic to a free semigroup on an alphabet.
**Proposition 13.2.** Suppose that \((S, \cdot)\) is a discrete positive semigroup with the property that every \(x \in S\) has a unique finite factoring over the set of irreducible elements \(I\). Then \((S, \cdot)\) is isomorphic to the free semigroup on \(I\).

**Proof.** Every \(x \in S\) has a finite factoring over \(I\), so that \(x = i_1 i_2 \cdots i_n\). If this factoring is unique, then clearly \((i_1, i_2, \ldots, i_n) \mapsto i_1 i_2 \cdots i_n\) is an isomorphism from the free semigroup \((I^*, \cdot)\) onto \((S, \cdot)\).

The left walk function, the left generating function, and the Möbius kernel follow from our previous work in Chapter 12.

**Proposition 13.3.** The left walk function \(\gamma_n\) of order \(n \in \mathbb{N}\) for \((S, \preceq)\) is given by

\[
\gamma_n(x) = \binom{d(x) + n}{n}, \quad x \in S
\]

**Exercise 13.1.** Find the left walk function of order \(n \in \mathbb{N}\) for the following graphs:

(a) \((S, \uparrow)\)
(b) \((S, \uparrow\uparrow)\)
(c) \((S, \prec)\)

**Proposition 13.4.** The left generating function \(\Gamma\) for \((S, \preceq)\) is given by

\[
\Gamma(x, t) = \frac{1}{(1 - t)^{d(x) + 1}}, \quad x \in S, \ |t| < 1
\]

**Exercise 13.2.** Find the left generating function for the following graphs:

(a) \((S, \uparrow)\)
(b) \((S, \uparrow\uparrow)\)
(c) \((S, \prec)\)

**Proposition 13.5.** The Möbius kernel \(M\) of \((S, \preceq)\) is given as follows: For \((x, y) \in S^2\), \(M(x, y) = 1\) if \(y = x\), \(M(x, y) = -1\) if \(y = xi\) for some \(i \in I\), and \(M(x, y) = 0\) otherwise.

The semigroup dimension is the number of letters.

**Proposition 13.6.** \(\dim(S, \cdot) = \#(I)\).

**Proof.** Note first that a homomorphism \(\phi\) from \((S, \cdot)\) into \((\mathbb{R}, +)\) can uniquely be specified by defining \(\phi(i)\) for all \(i \in I\) and then defining

\[
\phi(i_1 i_2 \cdots i_n) = \phi(i_1) + \phi(i_2) + \cdots + \phi(i_n)
\]

Thus, if \(\phi\) is a such a homomorphism and \(\phi(i) = 0\) for all \(i \in I\), then \(\phi(x) = 0\) for all \(x \in S\). Now suppose that \(B \subseteq S\) and \(\#(B) < \#(I)\). We will show that there exists a nonzero homomorphism from \((S, \cdot)\) into \((\mathbb{R}, +)\) with \(\phi(x) = 0\) for all \(x \in B\). Let \(I_B\) denote the set of letters contained in the words in \(B\). Suppose first that \(I_B\) if a proper subset of \(I\) (and note that this must be the case if \(I\) is infinite). Define a homomorphism
The set of linear, homogeneous equations

\[
\phi(i_j) + \phi(i_{j2}) + \cdots + \phi(i_{jm_j}) = 0, \quad j = 1, 2, \ldots, k
\]

has \(n\) unknowns, namely \(\phi(i)\) for \(i \in I\), but only \(k\) equations. Hence there exists a non-trivial solution. The homomorphism so constructed satisfies \(\phi(x) = 0\) for \(x \in B\). 

We can use the strict positive semigroup to provide a simple example of a discrete semigroup in which multiples of counting measure \(\#\) are not the only left invariant measures.

**Example 13.1.** Suppose that \(I\) has at least two elements and consider the strict positive semigroup \((S_+, \cdot)\) associated with the free semigroup \((S, \cdot)\). That is, \(S_+ = S - \{e\}\) while \(\cdot\) is still concatenation. Define the function \(g : S_+ \to (0, \infty)\) by \(g(x) = c_i\) if the terminal letter of \(x\) is \(i \in I\), where \(c_i \in (0, \infty)\) for each \(i \in I\) and \(c_i \neq c_j\) for some \(i, j \in I\). Let \(\lambda\) be the positive measure on \(S_+\) with density \(g\) (relative to \(\#\)). That is,

\[
\lambda(A) = \sum_{x \in A} g(x) = \sum_{i \in I} c_i n_i(A), \quad A \subseteq S_+
\]

where \(n_i(A)\) is the number of words in \(A\) with terminal letter \(i\) for each \(i \in I\). Then \(\lambda\) is left invariant: If \(x \in S\) and \(A \subseteq S\) then the words in \(xA\) have the same terminal letters as the corresponding words in \(A\), so \(\lambda(xA) = \lambda(A)\). But \(\lambda\) is clearly not a multiple of counting measure.

### 13.2 Probability

Suppose that \(X\) is a random variable with values in \(S\) and with density function \(f\). By definition, the right probability function \(F\) of \(X\) for \((S, \preceq)\) is given by

\[
F(x) = \mathbb{P}(X \geq x) = \sum_{y \preceq x} f(y) = \sum_{t \in S} f(xt), \quad x \in S
\]

The following several results follow from Chapter 12 on rooted trees:

**Proposition 13.7.** A function \(F : S \to [0, \infty]\) is a right probability function for \((S, \preceq)\) if and only if

(a) \(F(e) = 1\)

(b) \(\sum_{i \in I} F(x_i) \leq F(x)\) for \(x \in S\)

(c) \(\sum_{d(x) = n} F(x) \to 0\) as \(n \to \infty\)

In this case, the density function \(f\) is given by

\[
f(x) = F(x) - \sum_{i \in I} F(x_i), \quad x \in S
\]

**Proposition 13.8.** If \(X\) is a random variable with values in \(S\) then

\[
\sum_{x \in S} \binom{d(x) + n}{n} \mathbb{P}(X \geq x) = \mathbb{E} \left[ \binom{d(X) + n + 1}{n + 1} \right], \quad n \in \mathbb{N}
\]

**Exercise 13.3.** Give the standard moment result for each of the following graphs:

(a) \((S, \uparrow)\)

(b) \((S, \\uparrow)\)

(c) \((S, \prec)\)
Lemma 13.1. For $\alpha \in (0, 1)$, the following recursive scheme defines a probability density function $f$ on $S$:

1. $f(e) = \alpha$
2. If $f(x)$ is defined for some $x \in S$ then define $f(xi)$ for $i \in I$ arbitrarily, subject to the conditions
   - $(a)$ $f(xi) > 0$ for $i \in I$
   - $(b)$ $\sum_{i \in I} f(xi) = (1 - \alpha)f(x)$

Proposition 13.9. Random variable $X$ has constant rate $\alpha \in (0, 1)$ for $(S, \preceq)$ if and only if $X$ has a density function $f$ on the type constructed in Lemma 13.1, and in this case,

- $(a)$ $X$ has constant rate $1/(1 - \alpha)^n$ for $(S, \uparrow^n)$ for each $n \in \mathbb{N}$.
- $(b)$ $X$ has constant rate $1/(2 - \alpha)$ for $(S, \Downarrow)$.
- $(c)$ $X$ has constant rate $\alpha/(1 - \alpha)$ for $(S, \prec)$.
- $(d)$ $N = d(X)$ has constant rate $\alpha$ for $(\mathbb{N}, \leq)$.

Recall that part $(d)$ means that $N$ has the geometric distribution on $\mathbb{N}$ with success parameter $\alpha$. Since we have a semigroup, we are particularly interested in exponential distributions. Let $P_I$ denote the collection of probability density functions $p = (p_i : i \in I)$ that have support $I$, so that $p_i > 0$ for $i \in I$ and $\sum_{i \in I} p_i = 1$.

Theorem 13.1. Random variable $X$ has an exponential distribution on $(S, \cdot)$ if and only if $X$ has density function $f$ of the form

$$f(x) = \alpha(1 - \alpha)^{d(x)} \prod_{i \in I} p_i^{d_i(x)}, \quad x \in S$$

where $\alpha \in (0, 1)$ is the rate constant, and where $p = (p_i : i \in I) \in P_I$.

Proof. We apply Corollary 5.6. Let $F$ denote the right probability function of $X$ and let $F(i) = \beta_i$ for $i \in I$. Then $\beta_i > 0$ for each $i \in I$ and the memoryless property requires that

$$F(x) = \prod_{i \in I} \beta_i^{d_i(x)}, \quad x \in S$$

The constant rate property requires that $\sum_{x \in S} F(x) < \infty$ (and the reciprocal of this sum is the rate parameter). But from the multinomial theorem

$$\sum_{x \in S} F(x) = \sum_{n=0}^{\infty} \sum \left\{ C(n_i : i \in I) \prod_{i \in I} \beta_i^{n_i} : (n_i : i \in I) \in \mathbb{N}, \sum_{i \in I} n_i = n \right\} = \sum_{n=0}^{\infty} \beta^n$$

where $\beta = \sum_{i \in I} \beta_i$. Hence we must have $\beta \in (0, 1)$ and the rate constant is $1 - \beta$. So $X$ has density function $f$ given by

$$f(x) = (1 - \beta) \prod_{i \in I} \beta_i^{d_i(x)}, \quad x \in S$$

Finally, we re-define the parameters: let $\alpha = 1 - \beta$ and let $p_i = \beta_i/\beta$ for $i \in I$. Hence

$$f(x) = \alpha \prod_{i \in I} [p_i(1 - \alpha)]^{d_i(x)} = \alpha(1 - \alpha)^{d(x)} \prod_{i \in I} p_i^{d_i(x)}, \quad x \in S$$

In particular, in the free semigroup, every memoryless distribution has constant rate and hence is exponential. But of course that also follows from the general theory via Theorem 5.7. The converse is not true. The exponential distribution with parameters $\alpha \in (0, 1)$ and $(p_i : i \in I) \in P_I$ corresponds to choosing

$$f(xi) = (1 - \alpha)p_i f(x), \quad x \in S, i \in I$$

in Lemma 13.1. But of course the family of constant rate distributions in this lemma includes many others not of this type. Here is a simple example:
Example 13.2. Suppose that $I = \{0, 1\}$, so that $S$ consists of bit strings. Let $a > 0$, $b > 0$, $a \neq b$, and $a + b < 1$. Define $F : S \to (0, 1]$ as follows:

$$F(e) = 1, \quad F(0x) = a^{d_1(x)}b^{d_0(x)}, \quad F(1x) = a^{d_1(x)}b^{d_0(x)} + 1, \quad x \in S$$

Then $F$ is a right probability function relative to $(S, \preceq)$ for a distribution with constant rate $1 - a - b$ but this distribution is not exponential for $(S, \cdot)$ or even a mixture of exponential distributions.

There is a simple interpretation of the exponential distribution in terms of a random product of independent, identically distributed letters.

Corollary 13.1. Consider a sequence of independent variables $U = (U_1, U_2, \ldots)$ taking values in $I$ with common density function $p \in P_I$. Let $N$ be independent of $U$ and have a geometric distribution with success parameter $\alpha \in (0, 1)$. Let $X$ be the random variable in $S$ defined by $X = U_1U_2\cdots U_N$. Then $X$ has the exponential distribution with parameters $\alpha$ and $p$.

**Proof.** Let $x = i_1i_2\cdots i_n \in S$ where $n \in N$ and $(i_1, i_2, \ldots, i_n) \in I^n$. Then

$$\mathbb{P}(X = x) = \mathbb{P}(X = x, N = n) = \mathbb{P}(X = x \mid N = n)\mathbb{P}(N = n) = \mathbb{P}(U_1 = i_1, U_2 = i_2, \ldots, U_n = i_n)\alpha(1 - \alpha)^{n-1} = \mathbb{P}(U_1 = i_1)\mathbb{P}(U_2 = i_2)\cdots\mathbb{P}(U_n = i_n)\alpha(1 - \alpha)^n$$

It’s worth remembering again that the geometric distribution with success parameter $\alpha \in (0, 1)$ is the exponential distribution on $(N, +)$ with rate $\alpha$. The exponential distribution on $(S, \cdot)$ is clearly invariant under a permutation of the letters. If $X = U_1U_2\cdots U_N$ has the exponential distribution in Corollary 13.1 and if $(V_1, V_2, \ldots, V_N)$ is a permutation of $(U_1, U_2, \ldots, U_N)$ then $Y = V_1V_2\cdots V_N$ has the same exponential distribution.

Corollary 13.2. Suppose that $X$ has the exponential distribution on $(S, \cdot)$, with parameters $\alpha \in (0, 1)$ and $p = (p_i : i \in I) \in P_I$.

(a) For $n \in N$, the distribution of $(d_i(X) : i \in I)$ given $d(X) = n$ is multinomial with parameters $n$ and $p$:

$$\mathbb{P}[d_i(X) = n_i, i \in I] = \binom{n}{n_i, i \in I}p_i^{n_i}, \quad n_i \in N, i \in I; \sum_{i \in I} n_i = n$$

(b) The distribution of $d_i(X)$ is exponential on $(N, +)$ (that is, geometric) with rate parameter $\alpha/(\alpha + (1 - \alpha)p_i)$ for each $i \in I$.

Corollary 13.3. Suppose again that $X$ has the exponential distribution on $(S, \cdot)$, with parameters $\alpha \in (0, 1)$ and $p = (p_i : i \in I) \in P_I$. Random variable $X$ maximizes entropy over all random variables $Y \in S$ with

$$\mathbb{E}[d_i(Y)] = \mathbb{E}[d_i(X)] = \frac{1 - \alpha}{\alpha}p_i, \quad i \in I$$

**Theorem 13.2.** Suppose again that $X$ has the exponential distribution on $(S, \cdot)$ with parameters $\alpha \in (0, 1)$ and $p \in P_I$. Then $X$ has a compound Poisson distribution.

**Proof.** Recall from the Corollary 13.1 that $X$ can be decomposed as

$$X = U_1U_2\cdots U_N$$

where $U = (X_1, X_2, \ldots)$ are independent and identically distributed on the alphabet $I$, with probability density function $(p_i : i \in I)$, and where $N$ is independent of $U$ and has the geometric distribution on $\mathbb{N}$ with success parameter $\alpha$. (That is, the exponential distribution on $(N, +)$ with rate $\alpha$.) But from Chapter 11, $N$ has a compound Poisson distribution and can be written in the form

$$N = M_1 + M_2 + \cdots + M_K$$
where $M = (M_1, M_2, \ldots)$ are independent and identically distributed on $\mathbb{N}_+$ with the logarithmic distribution

$$
P(M = n) = \frac{(1 - \alpha)^n}{n \ln \alpha}, \quad n \in \mathbb{N}_+$$

and where $K$ is independent of $M$ and has the Poisson distribution with parameter $- \ln \alpha$. It follows that we can take $X, M$, and $K$ mutually independent, and thus

$$X = V_1 V_2 \cdots V_K$$

where $V_i = U_{M_{i-1}+1} \cdots U_{M_i}$. Note that $V = (V_1, V_2, \ldots)$ are independent and identically distributed on $S$ and thus $X$ has a compound Poisson distribution.

Example 13.3. The following list gives 10 simulated values of $X$, where $X$ has the exponential distribution on the free semigroup of bit strings $\{0,1\}^\ast$, with parameters $\alpha = 1/5$ and $p = (2/5, 3/5)$.

110100, 1111100000110, 111011, e, 100000110111100, 11011101, e, 000010, 011110, 0

From Corollary 13.1, it’s easy to construct the random walk on $(S, \cdot)$ corresponding to the exponential distribution with parameters $\alpha \in (0,1)$ and $p = (p_i : i \in I) \in P_I$. Let $U = (U_1, U_2, \ldots)$ be independent variables in $I$ with common density $p$ so that $P(U_i = i) = p_i$ for $i \in I$ and $j \in \mathbb{N}_+$. Let $J = (J_1, J_2, \ldots)$ be independent variables each with the geometric distribution on $\mathbb{N}$ with rate $\alpha$, and with $J$ independent of $U$. Thus $K_n = \sum_{i=1}^n J_i$ has the negative binomial distribution with parameters $\alpha$ and $n$ for $n \in \mathbb{N}_+$. Let $X_n = U_1 \cdots U_{K_n}$ for $n \in \mathbb{N}_+$. Then $X = (X_1, X_2, \ldots)$ is the random walk associated with the exponential distribution. For $n \in \mathbb{N}_+$, the probability density function $f_n$ of $X_n$ is given by

$$f_n(x) = \binom{d(x) + n - 1}{n - 1} \alpha^n (1 - \alpha)^{d(x)} \prod_{i \in I} p_i^{d(x)}, \quad x \in S$$

Of course, this also follows from the general theory. From our standard “most random” interpretation, observing $X_{n+1}$ gives no information about the increasing sequence of prefixes $(X_1, X_2, \ldots, X_n)$.

Corollary 13.4. Suppose that $X = (X_1, X_2, \ldots)$ is the random walk on $(S, \cdot)$ associated with the exponential distribution with parameters $\alpha \in (0,1)$ and $p = (p_i : i \in I) \in P_I$. Let $n \in \mathbb{N}_+$.

(a) The conditional distribution of $[d_i(X_n) : i \in I]$ given $d(X_n) = m \in \mathbb{N}$ is multinomial with parameters $m$ and $p$.

(b) The distribution of $d_i(X_n)$ is negative binomial with parameters $\alpha/\alpha + (1 - \alpha)p_i$ and $n$ for each $i \in I$.

Our final discussion concerns quotient spaces as studied in Chapter 9. The most natural way to form a sub-semigroup of $(S, \cdot)$ is to start with a proper subset of letters $J \subset I$ and then consider the set of finite words $T = J^\ast$ with letters in $J$. Clearly $(T, \cdot)$ is a complete sub-semigroup of $(S, \cdot)$, and both are free semigroups. The quotient space $S/T$ consists of the empty word $e$ and words $x \in S$ with first letter in $I \setminus J$. The assumptions in Chapter 9 are satisfied, so $x \in S$ has the unique factoring $x = yz$ with $y \in T$ and $z \in S/T$. Clearly $y$ is the largest initial prefix of $x$ with letters in $J$ and $z$ is the corresponding suffix. The following result follows easily from the general theory:

Theorem 13.3. Suppose that $X$ has the exponential distribution on $(S, \cdot)$ with parameters $\alpha \in (0,1)$ and $p = (p_i : i \in I) \in P_I$, and let $p_J = \sum_{j \in J} p_j$. Consider the factoring $X = YZ$ where $Y$ takes values in $S$ and $Z$ takes values in $S/T$. Then

(a) The distribution of $Y$ is the same as the conditional distribution of $X$ given $X \in T$. This distribution is exponential distribution on $(T, \cdot)$ with parameters $1 - (1 - \alpha)p_J$ and $q = (q_j : j \in J) \in P_J$ where $q_j = p_j/p_J$ for $j \in J$.

(b) The distribution of $Z$ is the same as the conditional distribution of $X$ given $X \in S/T$, with density function $h$ given by

$$h(x) = \frac{\alpha(1 - \alpha)^{d(x)} \prod_{i \in I} p_i^{d_i(x)}}{1 - (1 - \alpha p_J)}, \quad x \in S/T$$
(c) \( Y \) and \( Z \) are independent.

\textbf{Proof.} The results follow from the general theory in Chapter 9, except for the form of the density functions. We use the construction in Corollary 13.1.

(a) Conditioning on \( N \) we have

\[ P(X \in T) = \sum_{n=0}^{\infty} P(U_1 \cdots U_n \in T \mid N = n) \alpha (1 - \alpha)^n = \sum_{n=0}^{\infty} p_J^n \alpha (1 - \alpha)^n = \frac{\alpha}{1 - (1 - \alpha)p_J} \]

Hence the conditional distribution of \( X \) given \( X \in T \) has density \( g \) defined by

\[ g(x) = \frac{f(x)}{\mathbb{P}(X \in T)} = [1 - (1 - \alpha)p_J] (1 - \alpha)^{d(x)} \prod_{i \in I} p_i^{d_i(x)} \]

\[ = [1 - (1 - \alpha)p_J] (1 - \alpha)p_J^{d_J(x)} \prod_{j \in J} \left( \frac{p_j}{p_J} \right)^{d_j(x)} \quad , \quad x \in T \]

For the last equation, note that if \( x \in T \) then \( d_i(x) = 0 \) for \( i \in I \setminus J \).

(b) Note that \( X \notin S/T \) if and only if \( N > 0 \) and \( U_1 \in J \). Hence \( \mathbb{P}(X \in S/T) = 1 - (1 - \alpha)p_J \). So the conditional distribution of \( X \) given \( X \in S/T \) has density \( h \) defined by

\[ h(x) = \frac{f(x)}{\mathbb{P}(X \in S/T)} = \frac{\alpha (1 - \alpha)^{d(x)} \prod_{i \in I} p_i^{d_i(x)}}{1 - (1 - \alpha)p_J}, \quad x \in S/T \]

\( \square \)
Chapter 14

Arithmetic Semigroups

14.1 Basics

The positive semigroups studied in this chapter have some interesting connections to abstract analytic number theory \cite{22} and in the most important special case, to classical number theory \cite{28}. We start with a preliminary definition to help with notation, and then give the main definition for the objects that we will study.

**Definition 14.1.** Suppose that $I$ is a countable set. Define

\[ N_I = \{ (n_i : i \in I) : n_i \in \mathbb{N} \text{ for all } i \in I \text{ and } n_i = 0 \text{ for all but finitely many } i \in I \} \]

So if $I$ if finite then $N_I = \mathbb{N}^I$, the set of functions from $I$ into $\mathbb{N}$.

**Definition 14.2.** An arithmetic semigroup is a discrete, commutative, positive semigroup $(S, \cdot)$ with the property that there exists a subset $I \subset S$ of prime elements such that every $x \in S$ has a unique factoring of the form

\[ x = \prod_{i \in I} i^{n_i} \text{ for some } (n_i : i \in I) \in N_I \]

The term unique in the definition means unique up to the ordering of the factors. In the factorization, $i^0 = e$, the identity element, so the products are finite products, and hence the definition makes sense. We will generally use subscript notation for functions defined on $I$, the set of prime elements.

**Definition 14.3.** Suppose that $(S, \cdot)$ is an arithmetic semigroup with $I$ as the set of prime elements. For $x \in S$ and $i \in I$, let $d_i(x)$ denote the power of $i$ in the factorization of $x$, and let $d(x) = \sum_{i \in I} d_i(x)$. If $d_i(x) \in \mathbb{N}^+$ then $i$ is a prime factor of $x$.

So $d_i : S \to \mathbb{N}$ for $i \in I$ and is a bit like a logarithm function. Also $d : S \to \mathbb{N}$ since for $x \in S$, $d_i(x) = 0$ for all but finitely many $i \in I$. The prime elements are also the irreducible elements of $(S, \cdot)$. For the partial order $\preceq$ associated with an arithmetic semigroup $(S, \cdot)$, note that $x \preceq y$ if and only if $x$ divides $y$ in the sense of the factorization. That is, $x \preceq y$ if and only if $d_i(x) \leq d_i(y)$ for $i \in I$ and then

\[ x^{-1} y = \prod_{i \in I} i^{d_i(y) - d_i(x)} \]

**Example 14.1.** The canonical example of an arithmetic semigroup is $(\mathbb{N}^+, \cdot)$ where $\cdot$ is ordinary multiplication, and where $I = \{2, 3, 5, 7, \ldots\}$ is the set of prime numbers. For the associated partial order, $x \preceq y$ if and only $x$ divides $y$ in the ordinary arithmetic sense, and then $x^{-1} y$ is $y/x$.

The division relation on $\mathbb{N}^+$ is usually denoted $|$. But this notation is inadequate because it lacks direction, so we will use our generic partial order notation.

**Example 14.2.** An arithmetic semigroup $(S, \cdot)$ with a single prime element $i$ is isomorphic to the standard discrete semigroup $(\mathbb{N}, +)$ studied in Chapter 11, since $S = \{i^n : n \in \mathbb{N}\}$ and $i^m i^n = i^{m+n}$ for $m, n \in \mathbb{N}$.  

126
Example 14.3. Let \((I^*, \cdot)\) be the free semigroup associated with a countable collection of letters \(I\), as studied in Chapter 13. If we impose the equations \(ij = ji\) for \(i, j \in I\), then the new semigroup \((S, \cdot)\) is no longer free, but is the arithmetic semigroup with \(I\) as the set of prime elements.

Other examples of arithmetic semigroups occur in various algebraic and topological structures. Also, arithmetic semigroups occur naturally as sub-semigroups. Suppose that \((S, \cdot)\) is an arithmetic semigroup with \(I\) as the set of prime elements. If \(J\) is a proper subset of \(I\) then the sub-semigroup \((T, \cdot)\) of \((S, \cdot)\) generated by \(J\) is clearly also an arithmetic semigroup, with \(J\) as the set of prime elements. The corresponding quotient space as discussed in Chapter 9 is the arithmetic semigroup generated by \(I - J\), with \(I - J\) as the set of prime elements.

At this point in our study, all that matters in an arithmetic semigroup is the set of prime elements and the powers in the unique factoring of an element. Hence the following proposition is not surprising, and has already been hinted at several times, particularly in Example 14.2.

Proposition 14.1. Suppose that \((S, \cdot)\) is an arithmetic semigroup with set of prime elements \(I\). Then \((S, \cdot)\) is isomorphic to the positive semigroup \((\mathbb{N}_I, +)\) where + is component-wise addition. An isomorphism is \(x \mapsto (d_i(x) : i \in I)\).

Proof. If \(x \in S\), then \((d_i(x) : i \in I) \in \mathbb{N}_I\) by definition of the arithmetic semigroup. If \((n_i : i \in I) \in \mathbb{N}_I\) then \(\prod_{i \in I} i^{n_i} \in S\) by the closure property of \((S, \cdot)\), so the mapping is onto. The mapping is one-to-one by the uniqueness of the factorization. Finally note that \(d_i(xy) = d_i(x) + d_i(y)\) for \(x, y \in S\) and \(i \in I\).

For the remainder of this section, we assume that \((S, \cdot)\) is an arithmetic semigroup with identity \(e\) and with \(I\) as the set of prime elements. In addition to the partial order graph \((S, \preceq)\), we have the usual associated graphs: the strict partial order graph \((S, \prec)\), the covering graph \((S, \uparrow)\), and the reflexive completion \((S, \uparrow)\) of \((S, \uparrow)\). The covering relation is the relation associated with \(I\), so that \(x \uparrow y\) if and only if \(y = xi\) for some \(i \in I\).

The partial order graph \((S, \preceq)\) is uniform. That is, for \(x \in S\), all paths from \(e\) to \(x\) have the same length in the covering graph \((S, \uparrow)\), namely \(d(x)\). Also, \((S, \preceq)\) is a lattice. If \(x, y \in S\) then \(x \wedge y\) is the element defined by \(d_i(x \wedge y) = \min\{d_i(x), d_i(y)\}\) for \(i \in I\) and similarly \(x \vee y\) is the element defined by \(d_i(x \vee y) = \max\{d_i(x), d_i(y)\}\) for \(i \in I\). The following definition restates some common ideas in the language of number theory.

Definition 14.4. Suppose that \(x, y \in S\).

(a) If \(u \in S\) satisfies \(u \preceq x\) and \(u \preceq y\) then \(u\) is a common divisor of \(x\) and \(y\).

(b) \(x \wedge y\) is the greatest common divisor of \(x\) and \(y\), also denoted \(\gcd(x, y)\). That is, \(x \wedge y\) is a common divisor of \(x\) and \(y\), and \(u \preceq x \wedge y\) for every common divisor \(u\) of \(x\) and \(y\).

(c) \(x\) and \(y\) are relatively prime or coprime if \(x \wedge y = e\).

(d) If \(v \in S\) satisfies \(x \preceq v\) and \(y \preceq v\) then \(v\) is a common multiple of \(x\) and \(y\).

(e) \(x \vee y\) is the least common multiple of \(x\) and \(y\), also denoted \(\text{lcm}(x, y)\). That is, \(x \vee y\) is a common multiple of \(x\) and \(y\), and \(x \vee y \preceq v\) for every common multiple \(v\) of \(x\) and \(y\).

Equivalently \(x, y \in S\) are relatively prime if they have no common prime factors. That is, no \(i \in I\) satisfies \(d_i(x) > 0\) and \(d_i(y) > 0\).

Definition 14.5. A function \(a : S \to [0, \infty)\) is an arithmetic function for \((S, \cdot)\).

(a) \(a\) is additive if \(a(xy) = a(x) + a(y)\) for \(x, y \in S\).

(b) \(a\) is multiplicative if \(a(xy) = a(x) a(y)\) whenever \(x, y \in S\) are relatively prime.

(c) \(a\) is completely multiplicative if \(a(xy) = a(x) a(y)\) for all \(x, y \in S\).

More generally, arithmetic functions can take values in \(\mathbb{R}\) or even \(\mathbb{C}\), but we will only need nonnegative arithmetic functions. Note that the functions \(d_i\) for \(i \in I\) are additive. If \(a\) is multiplicative then

\[
a(x) = \prod_{i \in I} a\left(i^{d_i(x)}\right), \quad x \in S
\]
If $a$ is completely multiplicative then
\[ a(x) = \prod_{i \in I} [a(i)]^{d_i(x)}, \quad x \in S \]
The product of two multiplicative (completely multiplicative) functions is also multiplicative (completely multiplicative).

**Proposition 14.2.** The left walk function $\tau_n$ of order $n \in \mathbb{N}$ for the graph $(S, \preceq)$ is given by
\[ \tau_n(x) = \prod_{i \in I} \binom{d_i(x) + n}{n}, \quad x \in S \]
$\tau_n$ is multiplicative for $n \in \mathbb{N}$.

**Proof.** Recall that the left walk function of order $n \in \mathbb{N}$ for the standard partial order graph $(\mathbb{N}, \leq)$ is $k \mapsto \binom{k+n}{n}$. Hence the form of $\tau_n$ given above follows from the isomorphism between $(S, \cdot)$ and $(\mathbb{N}, +)$ in Proposition 14.1 and results on product spaces. If $x, y$ are relatively prime then no $i \in I$ satisfies $d_i(x) > 0$ and $d_i(y) > 0$ and therefore
\[ \tau_n(xy) = \prod_{i \in I} \binom{d_i(xy) + n}{n} = \prod_{i \in I} \binom{d_i(x) + d_i(y) + n}{n} = \prod_{i \in I, d_i(x) > 0} \binom{d_i(x) + n}{n} \prod_{i \in I, d_i(y) > 0} \binom{d_i(y) + n}{n} = \tau_n(x) \tau_n(y) \]
\[ \square \]
Recall that for $n \in \mathbb{N}$ and $x \in S$, $\tau_n(x)$ gives the number of $n+1$ factorings of $x$, an important function in analytic number theory. In particular, $\tau(x) = \# \{ u \in \mathbb{N}_+ : u \preceq x \}$ is the number of divisors of $x$.

**Exercise 14.1.** Find the left walk function of order $n \in \mathbb{N}$ for each of the following graphs:
(a) $(S, \prec)$
(b) $(S, \uparrow)$
(c) $(S, \Uparrow)$

**Problem 14.1.** Find a closed form expression for the left generating functions of $(S, \preceq)$, $(S, \prec)$, $(S, \uparrow)$, and $(S, \Uparrow)$.

The next results relate to the Möbius function of the semigroup $(S, \cdot)$ and the corresponding Mönchus kernel of the graph $(S, \preceq)$, and are well known in analytic number theory. Recall from Definition 3.11 that $\mu$ is defined inductively by $\mu(e) = 1$ and $\mu(x) = -\sum_{t \prec x} \mu(t)$.

**Proposition 14.3.** The Mönchus function $\mu$ of $(S, \cdot)$ is multiplicative.

**Proof.** We use induction on the well-founded partial order graph $(S, \preceq)$. Suppose that $x, y \in S$ are relatively prime, and that $\mu(uv) = \mu(u)\mu(v)$ for every $u, v \in S$ that are relatively prime with $uv \prec xy$. Our goal is to show that $\mu(xy) = \mu(x)\mu(y)$. Since $x, y$ are relatively prime, $t \prec xy$ if and only if $t = uv$ where $u \preceq x$, $v \preceq y$ and $uv \neq xy$. Moreover, the factoring is unique. Hence
\[ \mu(xy) = -\sum_{t \prec xy} \mu(t) = -\sum_{u \prec x} \sum_{v \prec y} \mu(uv) - \sum_{u \prec x} \mu(xv) - \sum_{u \prec x} \mu(uy) \]
\[ = -\sum_{u \prec x} \sum_{v \prec y} \mu(u)\mu(v) - \sum_{u \prec x} \mu(x)\mu(v) - \sum_{u \prec x} \mu(u)\mu(y) \]
\[ = -\sum_{u \prec x} \mu(u)\sum_{v \prec y} \mu(v) - \mu(x)\sum_{v \prec y} \mu(v) - \mu(y)\sum_{v \prec y} \mu(v) \]
\[ = -\mu(x)\mu(y) + \mu(x)\mu(y) + \mu(x)\mu(y) = \mu(x)\mu(y) \]
\[ \square \]
Proposition 14.4. The Möbius function \( \mu \) of \((S, \cdot)\) is given as follows:

(a) If \( x \) is a product of \( k \) distinct prime elements for some \( k \in \mathbb{N} \) then \( \mu(x) = (-1)^k \).
(b) If \( x \) is not the product of distinct primes then \( \mu(x) = 0 \).

Proof. Recall that an empty product is interpreted as \( e \).

(a) By definition, \( \mu(\epsilon) = 1 \), and \( \mu(i) = -\mu(\epsilon) = -1 \) if \( i \in I \). If \( x = i_1i_2\cdots i_k \) is a product of \( k \) distinct primes where \( k \in \{2, 3, \ldots\} \) then by Proposition 14.3,

\[
\mu(x) = \mu(i_1)\mu(i_2)\cdots\mu(i_k) = (-1)^k
\]

(b) Suppose first that \( i \in I \) and \( n \in \{2, 3, \ldots\} \). Then

\[
\mu(i^n) = -\sum_{k=0}^{n-1} \mu(i^k) = -\sum_{k=0}^{n-2} \mu(i^k) - \mu(i^{n-1}) = \mu(i^{n-1}) - \mu(i^{n-1}) = 0
\]

More generally, if \( x \in S \) is not the product of distinct primes, then \( x = \prod_{i \in I} i^{n_i} \) where \( n_i \in \{2, 3, \ldots\} \) for some \( i \in I \). By Proposition 14.3,

\[
\mu(x) = \prod_{i \in I} \mu(i^{n_i}) = 0
\]

\[\square\]

Definition 14.6. An element \( x \in S \) that is the product (possibly empty) of distinct primes is square free. That is, \( x \) is square free if \( d_i(x) \in \{0, 1\} \) for all \( i \in I \).

Corollary 14.1. The Möbius kernel \( M \) of \((S, \preceq)\) is given by \( M(x, y) = \mu(x^{-1}y) \) if \( x \preceq y \), and is 0 otherwise.

The dimension of the semigroup is the number of prime elements.

Proposition 14.5. \( \dim(S, \cdot) = \#(I) \).

Proof. By Corollary 3.4, \( \dim(S, \cdot) \leq \#(I) \), so suppose that \( C \subset S \) with \( \#(C) < \#(I) \). We need to show that \( C \) cannot be a critical set, that is, there exists a non-trivial homomorphism \( \varphi \) from \((S, \cdot)\) into \((\mathbb{R}, +)\) with \( \varphi(x) = 0 \) for every \( x \in C \). Let \( J \subseteq I \) denote the set of prime factors of the elements of \( C \). The condition that \( \varphi(x) = 0 \) for \( x \in S \) is equivalent to

\[
\sum_{j \in J} d_j(x)\varphi(j) = 0, \quad x \in C
\]

(14.1)

If \( J \subseteq I \), then define \( \varphi(j) = 0 \) for \( j \in J \) and \( \varphi(i) = 1 \) for \( i \in I - J \). If \( J = I \), then (14.1) is a set of \( \#(C) \) linear, homogeneous equations in \( \#(I) \) unknowns. Since \( \#(C) < \#(I) \), there exists a non-trivial solution \( \varphi(i) \) for \( i \in I \). In either case, we then extend \( \varphi \) to a non-trivial homomorphism from \((S, \cdot)\) to \((\mathbb{R}, +)\) by the rule

\[
\varphi(x) = \sum_{i \in I} d_i(x)\varphi(i), \quad x \in S
\]

\[\square\]

It is usually assumed that an arithmetic semigroup has a norm function in the following sense:

Definition 14.7. A norm \( |\cdot| \) for \((S, \cdot)\) is a function on \( S \) with the following properties:

(a) \( |e| = 1 \)
(b) \( |i| > 1 \) for \( i \in I \).
(c) \( |xy| = |x||y| \) for \( x, y \in S \)
(d) \( N(t) < \infty \) for \( t \in (0, \infty) \) where \( N(t) = \#\{x \in S : |x| \leq t\} \).
By (c), the norm is completely multiplicative so
\[ |x| = \prod_{i \in I} |i|^{|a(i)|}, \quad x \in S \]  
(14.2)

In particular from (b) \(|x| > 1\) if \(x \in S_+\). Conversely, it’s easy to construct a norm on an arithmetic semigroup: let \(|e| = 1\) and define \(|i| > 1\) arbitrarily for \(i \in I\). Then define \(|x|\) by (14.2) for \(x \in S_+ - I\).

**Example 14.4.** For the standard arithmetic semigroup \((\mathbb{N}_+, \cdot)\) in Example 14.1, the standard norm is the identity function on \(\mathbb{N}_+\) so that \(|x| = x\) for \(x \in \mathbb{N}_+\). Hence \(N(t) = \lfloor t \rfloor\) for \(t \in (0, \infty)\). The norm is an essential part of \((\mathbb{N}_+, \cdot)\) so that the actual positive integers, and not just the counts of the prime exponents, are part of the theory.

**Example 14.5.** The arithmetic semigroup \((S, \cdot)\) with a single prime element \(i\) in Example 14.2 is isomorphic to \((\mathbb{N}, +)\). Any function of the form \(i^n \mapsto c^n\) with \(c \in (1, \infty)\) is a norm for \((S, \cdot)\).

For the following definitions, we assume that the arithmetic semigroup \((S, \cdot)\) has a fixed norm.

**Definition 14.8.** Suppose that \(a\) is an arithmetic function. The corresponding Dirichlet series \(A\) is the formal series defined by by
\[ A(t) = \sum_{x \in S} \frac{a(x)}{|x|^t}, \quad t \in (0, \infty) \]
If the series converges for some \(t > 0\), then the series converges for \(t\) in an interval of the form \((t_0, \infty)\).

**Definition 14.9.** The Dirichlet series corresponding to the constant function \(1\) is the *zeta function*:
\[ \zeta(t) = \sum_{x \in S} \frac{1}{|x|^t}, \quad t \in (1, \infty) \]

**Example 14.6.** These definitions are most important for the standard arithmetic semigroup \((\mathbb{N}_+, \cdot)\) with the standard norm. In this case, the Dirichlet series \(A\) corresponding to an arithmetic function \(a\) is the standard Dirichlet series in number theory:
\[ A(t) = \sum_{x=1}^{\infty} \frac{a(x)}{x^t}, \quad t \in (t_0, \infty) \]
In particular, the zeta function is the standard Riemann zeta function
\[ \zeta(t) = \sum_{x=1}^{\infty} \frac{1}{x^t}, \quad t \in (1, \infty) \]
For \((\mathbb{N}_+, \cdot)\) there is a one-to-one correspondence between the arithmetic function \(a\) and the series function \(A\). Given \(a\), we compute \(A\), of course, as the series in the definition. Conversely, given \(A\) defined on the interval of convergence \((t_0, \infty)\), we can recover the arithmetic function \(a\) [16].

**Example 14.7.** An arithmetic semigroup \((S, \cdot)\) with a single prime element \(i\) can be identified with the standard discrete semigroup \((\mathbb{N}, +)\) as in Example 14.2. Suppose that the norm is given by \(|n| = c^n\) where \(c \in (1, \infty)\). The Dirichlet series \(A\) corresponding to the function \(a\) is given by
\[ A(t) = \sum_{n=0}^{\infty} \frac{a(n)}{c^{nt}} \]
In particular, the zeta function is given by
\[ \zeta(t) = \sum_{n=0}^{\infty} \frac{1}{c^{nt}} = \frac{c^t}{c^t - 1}, \quad t \in (0, \infty) \]
We will need one more special function.

**Definition 14.10.** The Mangoldt function \(\Lambda\) is the arithmetic function defined as follows:
\[ \Lambda(x) = \begin{cases} \ln |i| & \text{if } x = i^n \text{ for some } i \in I \text{ and } n \in \mathbb{N}_+ \\ 0 & \text{otherwise} \end{cases} \]
14.2 Probability

14.2.1 Distributions in General

We assume again that we have an arithmetic semigroup \((S, \cdot)\) with identity element \(e\) and with \(I\) as the set of prime elements. Also as before, \((S, \preceq)\) denotes the corresponding partial order graph. We also assume that we have a fixed norm \(|\cdot|\) for \((S, \cdot)\), but we will try to clearly distinguish the results that require the norm structure from those that do not.

If \(X\) is a random variable with values in \(S\) and with density function \(f\) then the right probability function of \(X\) for \((S, \preceq)\) is given by

\[
F(x) = \mathbb{P}(x \preceq X) = \sum_{x \leq y} f(y) = \sum_{z \in S} f(xz), \quad x \in S
\]

Under mild conditions, the right probability function determines the distribution. Let \(\mu\) denotes the Möbius function of \((S, \cdot)\)

**Corollary 14.2.** Suppose that \(F\) is a right probability function of a distribution for \((S, \preceq)\). If \(\sum_{x \in S} F(x) < \infty\) then \(F\) uniquely determines the distribution, and the probability density function \(f\) is given by

\[
f(x) = \sum_{x \preceq y} \mu(x^{-1}y)F(y) = \sum_{z \in S} \mu(z)F(xz)
\]

**Proof.** This follows immediately from Theorem 4.4. \(\square\)

Our next corollary is the standard moment result for \((S, \preceq)\). As before, let \(\tau_n\) denote the left walk function of order \(n \in \mathbb{N}\) for \((S, \preceq)\), which in this context is best thought of as the divisor function of order \(n\).

**Corollary 14.3.** Suppose that \(X\) is a random variable taking values in \(S\) and with right probability function \(F\) for \((S, \preceq)\). Then

\[
\sum_{x \in S} \tau_n(x)F(x) = \mathbb{E}[\tau_{n+1}(X)], \quad k \in \mathbb{N}
\]

In particular, \(\sum_{x \in S} F(x) = \mathbb{E}[\tau(X)]\), the expected number of divisors of \(X\).

In terms of the prime powers, note that if \(X\) is a random variable with values in \(S\) then

\[
X = \prod_{i \in I} i^{N_i}
\]

where \(N_i = d_i(X)\) takes values in \(\mathbb{N}\) for \(i \in I\) and with probability 1, \(N_i = 0\) for all but finitely many \(i \in I\). Let \(F\) denote the right probability function of \(X\) as before. then

\[
F(x) = \mathbb{P}(X \succeq x) = \mathbb{P}[N_i \geq d_i(x) \text{ for all } i \in I], \quad x \in S
\]

In particular, if \(F_i\) denotes the right probability function of \(N_i\) for the graph \((\mathbb{N}, \succeq)\) for then \(F(i^n) = F_i(n)\) for \(i \in I\) and \(n \in \mathbb{N}\). If \((N_i : i \in I)\) is an independent sequence then

\[
F(x) = \prod_{i \in I} F_i[d_i(x)], \quad x \in S
\]

Conversely, we can construct a random variable \(X\) in \(S\) from random prime powers. A simple result from analysis will be useful throughout this chapter.

**Proposition 14.6.** Suppose that \(p_i \in (0, 1)\) for \(i \in I\). Then

\[
\prod_{i \in I} p_i > 0 \text{ if and only if } \sum_{i \in I} (1 - p_i) < \infty
\]
Definition 14.11. Let \((0, 1)_I\) denote the set of all functions \(p = (p_i : i \in I)\) from \(I\) into \((0, 1)\) with the property that \(\prod_{i \in I} p_i > 0\), or equivalently \(\sum_{i \in I} (1 - p_i) < \infty\). If \(I\) is finite then \((0, 1)_I = (0, 1)^I\), the set of functions from \(I\) into \((0, 1)\).

Proposition 14.7. Suppose that \(N_i\) is a random variable with values in \(\mathbb{N}\) for \(i \in I\), and let \(p_i = \mathbb{P}(N_i = 0)\) for \(i \in I\). If \((p_i : i \in I) \in (0, 1)_I\) then \(X = \prod_{i \in I} i^{N_i}\) is a well-defined random variable with values in \(S\). If \((N_i : i \in I)\) is an independent sequence, the converse is true.

Proof. Note that \(p_i\) is the probability that \(i\) is not a prime factor of \(X\). By the first Borel-Cantelli Lemma, if

\[
\sum_{i \in I} (1 - p_i) = \sum_{i \in I} \mathbb{P}(N_i > 0) < \infty
\]  

(14.3)

then with probability 1, \(N_i = 0\) for all but finitely many \(i \in I\), and hence \(X\) is well defined. If \((N_i : i \in I)\) is a sequence of independent variables, then the converse is true by the second Borel-Cantelli Lemma. That is, if \(X = \prod_{i \in I} i^{N_i}\) is a well-defined random variable with values in \(S\) then with probability 1, \(N_i = 0\) for all but finitely many \(i\) and hence (14.3) holds. Note in this case that \(\mathbb{P}(X = e) = \prod_{i \in I} p_i\). \(\square\)

When the prime powers are independent, there is a simple moment result for a completely multiplicative function of \(X\) in terms of the probability generating functions.

Proposition 14.8. Suppose that \(X\) is a random variable with values in \(S\). Let \(P_i\) denote the probability generating function of \(N_i = d_i(X)\) for \(i \in I\). If \((N_i : i \in I)\) is an independent sequence and \(a\) is completely multiplicative then

\[
\mathbb{E}[a(X)] = \prod_{i \in I} P_i[a(i)]
\]

Proof. By independence and the completely multiplicative property,

\[
\mathbb{E}[a(X)] = \mathbb{E} \left[ \prod_{i \in I} a(i)^{N_i} \right] = \prod_{i \in I} \mathbb{E} \left[ a(i)^{N_i} \right] = \prod_{i \in I} P_i[a(i)]
\]

\(\square\)

With the norm structure in place, we can also define probability distributions in terms of Dirichlet series.

Definition 14.12. Suppose that \(a\) is a positive arithmetic function and that the corresponding Dirichlet series \(A\) converges on the interval \((t_0, \infty)\) for some \(t_0 \in (0, \infty)\). The Dirichlet distribution on \((S, \cdot)\) corresponding to \(A\) with parameter \(t \in (t_0, \infty)\) has probability density function \(f\) given by

\[
f(x) = \frac{a(x)}{|x|^t A(t)} , \quad x \in S
\]

So \(f\) is proportional to the function \(x \mapsto a(x)/|x|^t\), and of course the normalizing constant must then be \(A(t)\). The most famous Dirichlet distribution corresponds to \(a(x) = 1\) for all \(x \in S\).

Definition 14.13. The zeta distribution with parameter \(t \in (1, \infty)\) has probability density function \(f\) given by

\[
f(x) = \frac{1}{\zeta(t)} \frac{|x|^{-t}}{\zeta(t)} , \quad x \in \mathbb{N}_+
\]

For the standard arithmetic semigroup \((\mathbb{N}_+, \cdot)\) in Example 14.1, the Dirichlet distributions, and in particular the zeta distribution, are the standard ones.

14.2.2 Exponential Distributions

Of course, we are particularly interested in exponential distributions for \((S, \cdot)\). We will first characterize the exponential distributions in terms of the random prime powers, and then in terms of Dirichlet distributions.
**Theorem 14.1.** Random variable $X$ has an exponential distribution on $(S, \cdot)$ if and only if $X$ has a probability density function $f$ has the form

$$f(x) = \prod_{i \in I} p_i (1 - p_i)^{d_i(x)}, \quad x \in S$$

where $(p_i : i \in I) \in (0, 1)^I$. The rate constant is $\prod_{i \in I} p_i$.

**Proof.** From Theorem 5.6, $X$ has an exponential distribution on $(S, \cdot)$ if and only if the right probability function $F$ of $X$ satisfies

$$F(xy) = F(x)F(y), \quad x, y \in S \quad (14.4)$$

and then the rate constant is $\alpha$. The memoryless condition (14.4) means that $F$ is completely multiplicative. This condition holds if and only if

$$F(x) = \prod_{i \in I} [F(i)]^{d_i(x)}, \quad x \in S \quad (14.5)$$

Let $p_i = 1 - F(i)$ for $i \in I$, so that $p_i$ is the probability that $i$ is not a prime factor of $X$. From our support assumption, $p_i \in (0, 1)$ for each $i \in I$ and so $F(x) = \prod_{i \in I} (1 - p_i)^{d_i(x)}$ for $x \in S$. Let $n = (n_i : i \in I)$ denote an element of $\mathbb{N}_I$. Then

$$\sum_{x \in S} F(x) = \sum_{n \in \mathbb{N}_I} F\left(\prod_{i \in I} i^{n_i}\right) = \sum_{n \in \mathbb{N}_I} \prod_{i \in I} (1 - p_i)^{n_i}$$

$$= \prod_{i \in I} \sum_{n = 0}^{\infty} (1 - p_i)^n = \prod_{i \in I} \frac{1}{p_i}$$

So condition 14.5 is satisfied if and only if $\prod_{i \in I} p_i > 0$, in which case the density function $f$ is given by

$$f(x) = \prod_{i \in I} p_i (1 - p_i)^{d_i(x)}, \quad x \in S$$

The basic moment result simplifies as usual. If $X$ has the exponential distribution with parameter $p \in (0, 1)_I$ then

$$\mathbb{E}[\tau_n(X)] = \prod_{i \in I} \frac{1}{p_i^n}, \quad n \in \mathbb{N}$$

For the standard arithmetic semigroup $(\mathbb{N}_+, \cdot)$, the exponential property has the following interpretation: The conditional distribution of $X/x$ given that $x$ divides $X$ is the same as the distribution of $X$. Thus, knowledge of one divisor of $X$ does not help in finding other divisors of $X$, a property that may have some practical applications. The exponential distribution given in Theorem 14.1 corresponds to independent, geometric distributions on the prime exponents.

**Corollary 14.4.** Random variable $X$ has the exponential distribution on $(S, \cdot)$ with parameter $p \in (0, 1)_I$ if and only if $X$ can be written in the form

$$X = \prod_{i \in I} i^{N_i}$$

where $N = (N_i : i \in I)$ is a sequence of independent random variables and $N_i$ has the geometric distribution on $\mathbb{N}$ with success parameter $p_i$ for each $i \in I$.

That is, $N_i$ has the exponential distribution on $(\mathbb{N}, +)$ with rate $p_i$ for each $i \in I$. This characterization could also be obtained from Section 8.3 and the isomorphism between $(S, \cdot)$ and the semigroup $(\mathbb{N}_I, +)$ given in Proposition 14.1. Using the norm structure, our next characterization is in terms of Dirichlet distributions.
Theorem 14.2. Random variable $X$ has an exponential distribution on $(S, \cdot)$ if and only if $X$ has a Dirichlet distribution with a completely multiplicative function.

Proof. Suppose that $a$ is completely multiplicative and that $X$ has the Dirichlet distribution corresponding to $a$ with parameter $t$ in the interval of convergence of the Dirichlet series $A$. Then the right probability function $F$ of $X$ is given by

$$F(x) = \mathbb{P}(X \geq x) = \frac{\sum_{y \geq x} a(y)|y|^{-t}}{A(t)} = \frac{\sum_{z \in S} a(xz)|xz|^{-t}}{A(t)}$$

$$= \frac{a(x)}{A(t)} \sum_{z \in S} a(z)|z|^{-t}$$

$$= \frac{a(x)|x|^{-t}}{A(t)} A(t) = a(x)|x|^{-t}, \quad x \in S$$

Hence $X$ has constant rate $1/A(t)$. Also, $X$ is memoryless, since $a$ is completely multiplicative:

$$F(xy) = \mathbb{P}(X \geq xy) = a(xy)(|xy|)^{-t} = a(x)a(y)|x|^{-t}|y|^{-t} = \mathbb{P}(X \geq x)\mathbb{P}(X \geq y)$$

Therefore $X$ has an exponential distribution. The converse is trivially true. Suppose that $X$ has an exponential distribution with right probability function $F$. For fixed $t \in (0, \infty)$, let $a(x) = |x|^t F(x)$ for $x \in \mathbb{N}_+$, and let $A(s) = \sum_{x \in S} a(x)|x|^{-s}$. Then $a$ is a completely multiplicative function and $A$ is the corresponding Dirichlet series. Moreover, $t$ is in the interval of convergence since $\sum_{x \in S} a(x)/|x|^t = \sum_{x \in S} F(x) < \infty$. The probability density function $f$ of $X$ is given by

$$f(x) = \mathbb{P}(X = x) = \frac{a(x)|x|^{-t}}{A(t)}, \quad x \in S$$

and so $X$ has the Dirichlet distribution corresponding to $a$ with parameter $t$. Note that since $a$ is completely multiplicative, all members of this Dirichlet family are exponential, from the first part of the theorem. □

Thus, a Dirichlet distribution with completely multiplicative coefficient function (in particular, the zeta distribution) has the representation given in Corollary 14.4. The geometric parameters of the random prime powers are are related to the arithmetic function by

$$1 - p_i = \mathbb{P}(X \geq i) = \frac{a(i)}{|i|^t}, \quad i \in I$$

For the standard case $(\mathbb{N}_+, \cdot)$, this result was considered surprising by Lin and Hu [23], but is quite natural in the context of positive semigroups. As a corollary, we get a probabilistic proof of the product formula for the Dirichlet series of a completely multiplicative function.

Corollary 14.5. Suppose that $a$ is a positive, completely multiplicative function on $S$ and that the corresponding Dirichlet series converges on $(t_0, \infty)$ for some $t_0 \in [0, \infty)$. Then

$$A(t) = \prod_{i \in I} \frac{|i|^t}{|i|^t - a(i)}, \quad t \in (t_0, \infty)$$

In particular

$$\zeta(t) = \prod_{i \in I} \frac{|i|^t}{|i|^t - 1}, \quad t \in (1, \infty)$$

The standard moment results can be rephrased as follows: If $X$ has the Dirichlet distribution with completely multiplicative function $a$ and parameter $t$, then $\mathbb{E}[\tau_n(X)] = A^n(t)$ for $n \in \mathbb{N}$. In particular, if $X$ has the zeta distribution with parameter $t > 1$ then $\mathbb{E}[\tau_n(X)] = \zeta^n(t)$.

We give another representation in terms of independent Poisson variables. This was obtained in [23], but we give an alternate derivation based on the exponential distribution. We give the result first in terms of the parameters of the prime powers and then in terms of the Dirichlet series. The second version uses the Mangoldt function $\Lambda$ in Definition 14.10.
Theorem 14.3. Suppose that $X$ has the exponential distribution on $(S, \cdot)$ with parameter $p \in (0, 1)_I$. For $x \in S_+$, let $\lambda_x = (1 - p_i)^n/n$ if $x = i^n$ for some $i \in I$ and $n \in \mathbb{N}_+$, and let $\lambda_x = 0$ otherwise. Then $X$ can be written in the form

$$X = \prod_{x \in S_+} x^{V_x}$$

where $(V_x : x \in S_+)$ is a sequence of independent variables and $V_x$ has the Poisson distribution with parameter $\lambda_x$ for $x \in S_+$.

Proof. We start with the representation in Corollary 14.4. By Theorem 11.3, we can write the geometrically distributed prime exponents in the form

$$N_i = \sum_{n=1}^{\infty} nV_{in}, \quad i \in I$$

where $\{V_{in} : i \in I, n \in \mathbb{N}_+\}$ are independent and $V_{in}$ has the Poisson distribution with parameter $(1 - p_i)^n/n$ for $i \in I$ and $n \in \mathbb{N}_+$. Substituting we have

$$X = \prod_{i \in I} \prod_{n=1}^{\infty} t^{nV_{in}} = \prod_{i \in I} \prod_{n=1}^{\infty} (t^n)^{V_{in}}$$

Now, for $x \in S_+$, let $V_x = V_{in}$ if $x = i^n$ for some $i \in I$ and $n \in \mathbb{N}_+$, and let $V_x = 0$ otherwise. Then $(V_x : x \in S_+)$ is a sequence of independent variables, $V_x$ has the Poisson distribution with parameter $\lambda_x$ for $x \in S_+$, and

$$X = \prod_{x \in S_+} x^{V_x}$$

\[\square\]

Corollary 14.6. Suppose that $X$ has the Dirichlet distribution with completely multiplicative coefficient function $a$ and parameter $t$. Then

$$X = \prod_{x \in S_+} x^{V_x}$$

where $(V_x : x \in S_+)$ is a sequence of independent variables, and for $x \in S_+$, $V_x$ has the Poisson distribution with parameter

$$\lambda_x = \frac{a(x)\Lambda(x)}{x^t \ln |x|}$$

The exponential distribution on $(S, \cdot)$ is a compound Poisson distribution.

Theorem 14.4. Suppose that $X$ has the exponential distribution on $(S, \cdot)$ with parameter $p \in (0, 1)_I$. Then $X$ has a compound Poisson distribution. Specifically

$$X = V_1V_2 \cdots V_M$$

where $V = (V_1, V_2, \ldots)$ is a sequence of independent and identically distributed variables on the set of prime powers $\{i^n : i \in I, n \in \mathbb{N}_+\}$, with common probability density function

$$P(V = i^n) = \frac{(1 - p_i)^n}{n \sum_{j \in I} \ln p_j}, \quad i \in I, n \in \mathbb{N}_+$$

The random index $M$ is independent of $V$ and has the Poisson distribution with parameter $-\sum_{i \in I} \ln p_i$.

Proof. Consider the representation given in the theorem. For $k \in \mathbb{N}_+$ and $i \in I$, random variable $V_k$ is a power of $i$ with probability $\ln p_i / \sum_{j \in I} \ln p_j$, independently over $k$. It follows that the number of factors $M_i$ that are powers of $i$ has the Poisson distribution with parameter $-\ln p_i$. Moreover, for $k \in \mathbb{N}_+$ and $i \in I$, the conditional distribution of $V_k$ given that $V_k$ is a power of $i$ is the logarithmic distribution with parameter $p_i$:

$$P(V_k = i^n \mid V_k \text{ is a power of } i) = \frac{(1 - p_i)^n}{n \ln p_i}, \quad n \in \mathbb{N}_+ \quad (14.6)$$

By grouping the factors according to the prime powers, it follows that $X$ has the same distribution as $\prod_{i \in I} i^{N_i}$, where $N_i = U_{i,1} + U_{i,2} + \cdots + U_{i,M_i}$ for $i \in I$ with the following properties satisfied:
• $U_{i,k}$ has the logarithmic distribution with parameter $p_i$ for $i \in I$ and $k \in \mathbb{N}_+$, given in (14.6).

• $M_i$ has the Poisson distribution with parameter $-\ln p_i$ for $i \in I$.

• The random variables $\{U_{i,k}, M_i : i \in I, k \in \mathbb{N}_+\}$ are independent.

It follows that $U_i = (U_{i,1}, U_{i,2}, \ldots)$ is an independent sequence and that $M_i$ is independent of $U_i$. By the standard compound result in Theorem 11.2, $N_i$ has the geometric distribution with success parameter $p_i$ for $i \in I$ and $(N_i : i \in I)$ is an independent sequence. Hence $X$ has the exponential distribution with parameter $(p_i : i \in I)$.

**Corollary 14.7.** Suppose that $X$ has the Dirichlet distribution corresponding to a positive, completely multiplicative function $a$ and Dirichlet series $A$, with parameter $t$ in the interval of convergence. Then $X$ has a compound Poisson distribution. Specifically

$$X = V_1 V_2 \cdots V_M$$

where $V = (V_1, V_2, \ldots)$ is a sequence of independent and identically distributed variables on the set of prime powers $\{i^n : i \in I, n \in \mathbb{N}_+\}$, with common probability density function

$$\mathbb{P}(V = i^n) = \frac{a^n(i)}{n^{|i|n} \ln A(t)}, \quad i \in I, n \in \mathbb{N}_+$$

The random index $M$ is independent of $V$ and has the Poisson distribution with parameter $\ln A(t)$.

**Corollary 14.8.** Suppose again that $X$ has the exponential distribution on $(S, \cdot)$ with parameter $p \in (0, 1)$. Then $X$ maximizes entropy over all random variables $Y$ in $S$ with $(M_i = d_i(Y) : i \in I)$ independent and $\mathbb{E}(M_i) = (1 - p_i)/p_i$ for each $i \in I$, The maximum entropy is

$$H(X) = -\sum_{i \in I} \left[ \ln(p_i) + \ln(1 - p_i) \frac{1 - p_i}{p_i} \right]$$

**Corollary 14.9.** $X$ maximizes entropy over all random variables $Y \in S$ with $\mathbb{E}(\ln|Y|) = \mathbb{E}(\ln|X|)$ and $\mathbb{E}[\ln a(Y)] = \mathbb{E}[\ln a(X)]$.

From [16],

$$\mathbb{E}(\ln|X|) = \frac{1}{A(t)} \sum_{x \in S} \ln(|x|)a(x)|x|^{-t} = \frac{A'(t)}{A(t)}$$

**Corollary 14.10.** Suppose that $k \in (0, \infty)$ and that $p_i > 1 - 1/|i|^k$ for $i \in I$. Then

$$\mathbb{E}(|X|^k) = \prod_{i \in I} \frac{p_i}{1 - |i|^k(1 - p_i)}$$

**Proof.** This follows from the result above since $x \mapsto |x|^k$ is completely multiplicative, and the probability generating function of the geometric distribution with success parameter $p \in (0, 1)$ is $t \mapsto p/[1 - t(1 - p)]$ for $|t| < 1/(1 - p)$.

**Corollary 14.11.** If $k < t - t_0$ then

$$\mathbb{E}(|X|^k) = \frac{A(t - k)}{A(t)}$$

**Proof.** In terms of the Dirichlet formulation, the density function $f$ of $X$ is given by $f(x) = a(x)/|x|^t A(t)$ for $x \in S$. Hence

$$\mathbb{E}(|X|^k) = \sum_{x \in S} |x|^k \frac{a(x)}{|x|^t A(t)} = \frac{1}{A(t)} \sum_{x \in S} \frac{a(x)}{|x|^{t-k}} = \frac{A(t-k)}{A(t)}$$

assuming that $t - k > t_0$. The result also follows from Proposition 14.10 and the product expansion of the Dirichlet series $A$ since $p_i = 1 - a(i)/|i|^t$ for $i \in I$. 

---

**CHAPTER 14. ARITHMETIC SEMIGROPS**

136
Next we obtain a result from [23]. Our proof is better because it takes advantage of the general theory of positive semigroups. Let \( L \) denote the adjacency kernel of \((S, \cdot)\). Suppose that \( a : \mathbb{N}_+ \to [0, \infty) \) is a nonnegative arithmetic function, not identically zero, and let \( A \) be the corresponding Dirichlet series with interval of convergence \((t_0, \infty)\).

**Theorem 14.5.** If \( X \) has the zeta distribution with parameter \( t > \max\{t_0, 1\} \). Then
\[
\mathbb{E}[(aL)(X)] = A(t)
\]

**Proof.** It follows immediately from the basic moment result that
\[
\mathbb{E}[(aL)(X)] = \zeta(t) \mathbb{E}[a(X)]
\]
since \( 1/\zeta(t) \) is the rate constant of the exponential distribution of \( X \). But
\[
\zeta(t) \mathbb{E}[a(X)] = \zeta(t) \sum_{x \in S} \frac{a(x)}{|x|^t} \zeta(t) = \sum_{x \in S} \frac{a(x)}{|x|^t} = A(t)
\]

14.2.3 Random Walks

Suppose now that \( U = (U_1, U_2, \ldots) \) is a sequence of independent variables, each with the exponential distribution on \((S, \cdot)\) with parameters \( p \in (0, 1) \). Let \( X_n = U_1 \cdots U_n \) for \( n \in \mathbb{N}_+ \) so that \( X = (X_1, X_2, \ldots) \) is the random walk on \((S, \cdot)\) associated with the exponential distribution.

**Corollary 14.12.** For \( n \in \mathbb{N} \),
\[
X_n = \prod_{i \in I} ^{x_{M_{ni}}} , \quad n \in \mathbb{N}_+
\]
where \( M_n = \{M_{ni} : i \in I\} \) is an independent sequence of variables and \( M_{ni} \) has the negative binomial distribution on \( \mathbb{N} \) with parameters \( n \) and \( p_i \) for \( i \in I \).

Hence \( X_n \) has density function \( f_n \) given by
\[
f_n(x) = \prod_{i \in I} ^{x_{M_{ni}}} \left( n + d_i(x) - 1 \right)^{d_i(x) - 1} (1 - p_i)^{d_i(x)} p_i^n , \quad x \in S
\]

Suppose now that \( U = (U_1, U_2, \ldots) \) is an independent sequence of random variables, each with the Dirichlet distribution corresponding to the completely multiplicative function \( a \), Dirichlet series \( A \), and parameter \( t \). Let \( X_n = U_1 U_2 \cdots U_n \) for \( n \in \mathbb{N}_+ \) so that \( X = (X_1, X_2, \ldots) \) is the corresponding random walk on \((S, \cdot)\).

**Corollary 14.13.** For \( n \in \mathbb{N} \), \( X_n \) has density function \( f_n \) given by
\[
f_n(x) = \frac{\tau_{n-1}(x) a(x)}{A^n(t)|x|^t} , \quad x \in S
\]

Thus, \( X_n \) also has a Dirichlet distribution for \( n \in \mathbb{N} \), but corresponding to a multiplicative function instead of a completely multiplicative coefficient function. It follows that
\[
\sum_{x \in S} \frac{\tau_{n-1}(x) a(x)}{|x|^t} = A^n(t) , \quad n \in \mathbb{N}_+
\]

In the special case of the zeta distribution with parameter \( t > 0 \), the density function \( f_n \) of \( X_n \) is given by
\[
f_n(x) = \frac{\tau_{n-1}(x)}{\zeta^n(t)|x|^t} , \quad x \in S
\]
for \( n \in \mathbb{N}_+ \), so in this special case it follows that
\[
\sum_{x \in S} \frac{\tau_{n-1}(x)}{|x|^t} = \zeta^n(t) , \quad n \in \mathbb{N}_+
\]

In both formulations, the density function of \( X_n \) is a special case of the general theory. That is, for \( n \in \mathbb{N}_+ \) and \( x \in S \), \( f_n(x) \) is the product of the \( n \)th power of the rate constant, the left walk function of order \( n - 1 \) at \( x \), and the right probability function at \( x \).
14.2.4 Related Graphs

The exponential distribution on \((S, \cdot)\) has constant rate not only for \((S, \preceq)\) but also for the other graphs naturally associated with \((S, \preceq)\). We have seen this type of result before, in Chapter 12 on rooted trees and again in Chapter 13 on the free semigroup.

**Theorem 14.6.** Suppose that \(X\) has the exponential distribution on \((S, \cdot)\) with parameter \(p \in (0, 1)\).

(a) For the graph \((S, \prec)\), \(X\) has right probability function \(F_1\) given by

\[
F_1(x) = \left[1 - \prod_{i \in I} p_i \right] \prod_{i \in I} (1 - p_i)^{n_i(x)}, \quad x \in S
\]

and has constant rate \(\prod_{i \in I} p_i / (1 - \prod_{i \in I} p_i)\).

(b) For the graph \((S, \uparrow)\), \(X\) has right probability function \(F_2\) given by

\[
F_2(x) = \sum_{j \in I} (1 - p_j) \left[ \prod_{i \in I - \{j\}} p_i (1 - p_i)^{n_i(x)} \right], \quad x \in S
\]

and has constant rate \(1 / \sum_{j \in I} (1 - p_j)\).

(c) For the graph \((S, \uparrow)\), \(X\) has right probability function \(F_3\) given by

\[
F_3(x) = \left[1 + \sum_{j \in I} (1 - p_j) \right] \left[ \prod_{i \in I} p_i (1 - p_i)^{n_i(x)} \right], \quad x \in S
\]

and has constant rate \(1 / \left[1 + \sum_{j \in I} (1 - p_j)\right]\).

**Proof.** Recall that \(X\) has right probability function \(F\) for \((S, \preceq)\) given by \(F(x) = \prod_{i \in I} (1 - p_i)^{n_i(x)}\) for \(x \in S\) and has constant rate \(\prod_{i \in I} p_i\) for the graph.

(a) This follows from results on reflexive completion.

(b) For the covering graph \((S, \uparrow)\),

\[
F_2(x) = \sum_{j \in I} f(x_j) = \sum_{j \in I} \left[ \prod_{i \in I - \{j\}} p_i (1 - p_i)^{n_i(x)} \right] p_j (1 - p_j)^{n_j(x) + 1}
\]

\[
= \sum_{j \in I} \left[ \prod_{i \in I - \{j\}} p_i (1 - p_i)^{n_i(x)} \right] (1 - p_j) = f(x) \sum_{j \in I} (1 - p_j), \quad x \in S
\]

Note that \(\sum_{j \in I} (1 - p_j) < \infty\) since \(\prod_{i \in I} p_i > 0\).

(c) This follows from (b) and results on reflexive completion.

\[\square\]

**Corollary 14.14.** Suppose that \(X\) has the Dirichlet distribution on \((S, \cdot)\) corresponding to the completely multiplicative function \(a\) and parameter \(t\).

(a) For the graph \((S, \prec)\), \(X\) has right probability function \(F_1\) given by

\[
F_1(x) = \left[1 - \frac{1}{A(t)} \right] \frac{a(x)}{|x|^t}, \quad x \in S
\]

and has constant rate \(1/[A(t) - 1]\).
(b) For the graph \((S, \uparrow)\), \(X\) has right probability function \(F_2\) given by

\[
F_2(x) = \frac{a(x)|x|^{-t}}{A(t)} \sum_{j \in I} a(j)/|j|^t, \quad x \in S
\]

and has constant rate \(1/\sum_{j \in I} a(j)/|j|^t\).

(c) For the graph \((S, \uparrow\uparrow)\), \(X\) has right probability function \(F_3\) given by

\[
F_3(x) = \frac{a(x)|x|^{-t}}{A(t)} \left[ 1 + \sum_{j \in I} a(j)/|j|^t \right], \quad x \in S
\]

and has constant rate \(1/\left[ 1 + \sum_{j \in I} a(j) \right]\).

Other constant rate distributions can be obtained for any of the graphs by mixing two or more constant rate distributions of the types given in the theorem, with the same rate.

**Problem 14.2.** Characterize all constant rate distributions for each of the four graphs.
Chapter 15

Paths and Grids

15.1 Preliminaries

In this chapter we study the (half) infinite path, finite paths, and grids. Although these are among the simplest graphs, they are surprisingly interesting in the context of our theory. A family of polynomials will play a central role.

Definition 15.1. Define the family of polynomials \( \{P_n : n \in \mathbb{N}\} \) by the following initial conditions and recursion relation:

\[
P_0(t) = 1, \quad P_1(t) = t, \quad t \in \mathbb{R} \quad (15.1)
\]

\[
P_{n+1}(t) = tP_n(t) - P_{n-1}(t), \quad n \in \mathbb{N}_+, \quad t \in \mathbb{R} \quad (15.2)
\]

The first few members of the family are

\[
P_0(t) = 1
\]

\[
P_1(t) = t
\]

\[
P_2(t) = t^2 - 1
\]

\[
P_3(t) = t^3 - 2t
\]

\[
P_4(t) = t^4 - 3t^2 + 1
\]

This family of polynomials is closely related to a family of Chebyshev polynomials, and with this connection, the important properties follow easily.

Lemma 15.1. \( P_n(t) = U_n(t/2) \) for \( n \in \mathbb{N} \) and \( t \in \mathbb{R} \) where \( U_n \) is the Chebyshev polynomial of the second kind with degree \( n \).

Proof. This follows directly from the defining conditions for \( \{U_n : n \in \mathbb{N}\} \):

\[
U_0(t) = 1, \quad U_1(t) = 2t, \quad t \in \mathbb{R}
\]

\[
U_{n+1}(t) = 2tU_n(t) - U_{n-1}(t), \quad n \in \mathbb{N}_+, \quad t \in \mathbb{R}
\]

\( \square \)

Theorem 15.1. For \( n \in \mathbb{N}_+ \), the roots of \( P_n \) are

\[
2 \cos \left( \frac{k}{n+1} \pi \right), \quad k \in \{1, 2, \ldots, n\}
\]

Proof. This follows directly from Lemma 15.1 and the roots of the Chebyshev polynomials. \( \square \)

Note that \( P_n \) has \( n \) distinct (and hence simple) roots in the interval \((-2, 2)\).

Theorem 15.2. For \( n \in \mathbb{N} \) and \( \theta \in (0, \pi) \),

\[
P_n(2 \cos \theta) = \frac{\sin[(n+1)\theta]}{\sin \theta}
\]

Proof. This follows from Lemma 15.1 and the trigonometric definition of the Chebyshev polynomials. \( \square \)
15.2 The Infinite Path

Consider the discrete graph $\langle \mathbb{N}, \rightarrow \rangle$ where $\rightarrow$ is the relation on $\mathbb{N}$ defined by $x \rightarrow x+1$ for $x \in \mathbb{N}$ and $x \rightarrow x-1$ for $x \in \mathbb{N}^+$. So $\langle \mathbb{N}, \rightarrow \rangle$ is simply the undirected path $(0, 1, 2, \ldots)$, clearly a basic and interesting algebraic structure.

**Theorem 15.3.** The left walk function for $\langle \mathbb{N}, \rightarrow \rangle$ is given recursively as follows.

$$\gamma_n(0) = \left(\frac{n}{\lfloor n/2 \rfloor}\right), \quad n \in \mathbb{N} \tag{15.3}$$

$$\gamma_n(1) = \gamma_{n+1}(0) = \left(\frac{n+1}{\lfloor (n+1)/2 \rfloor}\right), \quad n \in \mathbb{N} \tag{15.4}$$

$$\gamma_n(x+1) = \gamma_{n+1}(x) - \gamma_n(x-1), \quad x \in \mathbb{N}^+, \ n \in \mathbb{N} \tag{15.5}$$

**Proof.** The proofs involve straightforward counting.

(a) Consider walks ending in 0. A walk of even length $2n$ must start in a state in $\{0, 2, \ldots, 2n\}$. A walk of odd length $2n+1$ must start in a state in $\{1, 3, \ldots, 2n+1\}$. For $n \in \mathbb{N}^+$, let $a_n(k)$ be the number of walks of length $2n$ starting in $k \in \{0, 2, \ldots, 2n\}$ and ending in 0, and let $b_n(k)$ be the number of walks of length $2n+1$ starting in $k \in \{1, 3, \ldots, 2n+1\}$ and ending in 0. So

$$\gamma_{2n}(0) = \sum_{j=0}^{n} a_n(2j), \quad \gamma_{2n+1}(0) = \sum_{j=0}^{n} b_n(2j+1); \quad n \in \mathbb{N}^+$$

Trivially $a_n(2n) = b_n(2n+1) = 1$ for $n \in \mathbb{N}^+$. Also, $b_n(k) = a_n(k-1) + a_n(k+1)$ for $k \in \{1, 3, \ldots, 2n-1\}$. Similarly, $a_{n+1}(0) = a_n(1)$ and $a_{n+1}(k) = b_n(k-1) + b_n(k+1)$ for $k \in \{2, 4, \ldots, 2n\}$. It then follows that $\gamma_{2n}(0) = 2\gamma_{2n-1}(0)$ and $\gamma_{2n+1}(0) = 2\gamma_{2n}(0) - a_n(0)$. This leads to (15.3).

(b) By definition, $\gamma_{n+1}(x) = (\gamma_n K)(x)$ for $n \in \mathbb{N}$, where as usual, $K$ is the adjacency matrix. With $x = 0$ we have $\gamma_{n+1}(0) = (\gamma_n K)(0) = \gamma_n(1)$ which gives (15.4).

(c) Finally, for $x \in \mathbb{N}^+$ we have $\gamma_{n+1}(x) = (\gamma_n K)(x) = \gamma_n(x-1) + \gamma_n(x+1)$. Solving for $\gamma_n(x+1)$ gives (15.5).

\[ \Box \]

**Theorem 15.4.** The left generating function $\Gamma(x, t)$ for $\langle \mathbb{N}, \rightarrow \rangle$ satisfies

$$\Gamma(0, t) = 1 + t\Gamma(1, t), \quad |t| < \frac{1}{2}$$

$$\Gamma(x, t) = 1 + t\Gamma(x+1, t) + t\Gamma(x-1, t), \quad x \in \mathbb{N}^+, \ |t| < \frac{1}{2}$$

**Proof.** First, it’s easy to see that for $x \in \mathbb{N}$, the generating function converges for $|t| < \frac{1}{2}$. From (15.4) we have $t^{n+1}\gamma_n(1) = t^{n+1}\gamma_{n+1}(0)$ for $|t| < \frac{1}{2}$ and $n \in \mathbb{N}$. Summing over $n \in \mathbb{N}$ gives

$$t\Gamma(1, t) = \Gamma(0, t) - \gamma_0(0) = \Gamma(0, t) - 1$$

From (15.5) we have

$$t^{n+1}\gamma_n(x+1) = t^{n+1}\gamma_{n+1}(x) - t^{n+1}\Gamma_n(x-1), \quad n \in \mathbb{N}, x \in \mathbb{N}^+, \ |t| < \frac{1}{2}$$

Summing over $n \in \mathbb{N}$ gives

$$t\Gamma(x+1, t) = \Gamma(x, t) - \gamma_0(x) - t\Gamma(x-1, t) = \Gamma(x, t) - 1 - t\Gamma(x-1, t), \quad x \in \mathbb{N}^+, \ |t| < \frac{1}{2}$$

\[ \Box \]

For the following result, $K$ denotes the kernel of $\langle \mathbb{N}, \rightarrow \rangle$, as usual, and $P_n$ denotes the polynomial in Definition 15.1 of degree $n \in \mathbb{N}$. 

**Theorem 15.5.** For $\beta \in \mathbb{R}$, the solution of the equation $Kg = \beta g$ is given by $g(x) = g(0)P_x(\beta)$ for $x \in \mathbb{N}$.

**Proof.** For $\beta \in \mathbb{R}$, then the equation $Kg = \beta g$ becomes

$$
g(1) = \beta g(0)$$
$$g(x + 1) = \beta g(x) - g(x - 1), \quad x \in \mathbb{N}_+
$$

Hence $g(x) = g(0)P_x(\beta)$ for $x \in \mathbb{N}$ by definition of the polynomials.

So on the vector space of all functions from $\mathbb{N}$ to $\mathbb{R}$, every $\beta \in \mathbb{R}$ is an (right) eigenvalue of $(\mathbb{N}, \to)$, and a corresponding eigenfunction is given by $x \mapsto P_x(\beta)$. But usually we are interested in eigenvalues and eigenfunctions of $(\mathbb{N}, \to)$ for the space $\mathcal{Z}_1$, where probability density functions live.

Suppose now that $X$ is a random variable with values in $\mathbb{N}$ and with density function $f$. The right probability function $F$ of $X$ for $(\mathbb{N}, \to)$ is given by

$$F(0) = f(1), \quad F(x) = f(x - 1) + f(x + 1), \quad x \in \mathbb{N}_+
$$

As usual, we have our basic support assumption that $F(x) > 0$ for all $x \in \mathbb{N}_+$. So in particular, $f(1) > 0$, and for every $x \in \mathbb{N}_+$, either $f(x - 1) > 0$ or $f(x + 1) > 0$. The right probability function does in fact determines the distribution, and we can characterize right probability functions:

**Theorem 15.6.** Suppose that $f$ is a density function on $\mathbb{N}$ and that $F$ is the corresponding right probability function for $(\mathbb{N}, \to)$. Then

$$f(0) = 2 - \sum_{x=0}^{\infty} F(x) \tag{15.6}$$
$$f(2x + 1) = (-1)^x \sum_{k=0}^{x} (-1)^k F(2k), \quad x \in \mathbb{N} \tag{15.7}$$
$$f(2x) = (-1)^x f(0) + (-1)^x \sum_{k=0}^{x} (-1)^k F(2k - 1), \quad x \in \mathbb{N}_+ \tag{15.8}
$$

Conversely, if $F$ is a positive function on $\mathbb{N}$ such that $1 < \sum_{x=0}^{\infty} F(x) < 2$ and such that the right side of (15.7) is positive for $x \in \mathbb{N}$ and the right side of (15.8) is positive for $x \in \mathbb{N}_+$, then $F$ is the right probability function for $(\mathbb{N}, \to)$ of a distribution on $\mathbb{N}$ with probability density function given by (15.6), (15.7), (15.8).

**Proof.** Suppose that $f$ is a density function on $\mathbb{N}$ and that $F$ is the right probability function of $f$ for $(\mathbb{N}, \to)$. Note that

$$\sum_{x=0}^{\infty} F(x) = f(1) + \sum_{x=1}^{\infty} [f(x - 1) + f(x + 1)] = f(0) + 2 \sum_{x=1}^{\infty} f(x) = f(0) + 2[1 - f(0)] = 2 - f(0)
$$

and hence (15.6) holds. For the odd-order terms, note first that $f(1) = F(0)$. Next, $F(2) = f(1) + f(3)$ so $f(3) = F(2) - F(0)$. Next, $F(4) = f(3) + f(5)$ so

$$f(5) = F(4) - f(3) = F(4) - F(2) + F(0)
$$

Continuing in this way gives (15.7). Similarly, $F(1) = f(0) + f(2)$, so $f(2) = F(1) - f(0)$, where $f(0)$ is given by (15.6). Next, $F(3) = f(2) + f(4)$ so

$$f(4) = F(3) - f(2) = F(3) - F(1) + f(0)
$$

Continuing in this way gives (15.8).

Conversely, suppose that $F$ is a positive function on $\mathbb{N}$ such that $1 < \sum_{x=0}^{\infty} F(x) < 2$ and with the right sides of (15.7) and (15.8) positive. Define $f$ by (15.6), (15.7), and (15.8). Then $f(x) \geq 0$ for $x \in \mathbb{N}$ and it’s easy to see from the definitions that $F(0) = f(1)$ and $F(x) = f(x - 1) + f(x + 1)$ for $x \in \mathbb{N}_+$. So we have

$$\sum_{x=0}^{\infty} F(x) = f(0) + 2 \sum_{x=1}^{\infty} f(x)$$
Therefore
\[ f(0) + \sum_{x=0}^{\infty} F(x) = 2 \sum_{x=0}^{\infty} f(x) \]
But by definition of \( f(0) \), the left side is 2. It follows that \( f \) is a density function on \( \mathbb{N} \) and \( F \) is the right probability function of \( f \) for \((\mathbb{N}, \rightarrow)\).

Here is our main result:

**Theorem 15.7.** The graph \((\mathbb{N}, \rightarrow)\) does not have a constant rate distribution. However, the geometric distribution on \( \mathbb{N} \) has constant rate on \( \mathbb{N}_+ \), and is the only such distribution.

**Proof.** Suppose that \( X \) is a random variable with value in \( \mathbb{N} \) having constant rate \( \alpha \). From our support assumptions, \( P(X = 0) > 0 \) and \( P(X \in \mathbb{N}_+) > 0 \). It follows that \( \mathbb{E}[\gamma(X)] \in (1, 2) \) where as usual, \( \gamma \) is the left walk function. But \( \mathbb{E}[\gamma(X)] = 1/\alpha \) from Theorem 4.8 so \( \alpha \in (\frac{1}{2}, 1) \).

Suppose next that \( X \) has probability density function \( f \). By the constant rate property, \( Kf = \frac{1}{\alpha} f \), so that
\[
\begin{align*}
  f(1) &= \frac{1}{\alpha} f(0) \\
  f(x + 1) &= \frac{1}{\alpha} f(x) - f(x-1), \quad x \in \mathbb{N}_+ 
\end{align*}
\]
The characteristic equation of (15.10) is \( r^2 - \frac{1}{\alpha} r + 1 = 0 \) which has roots
\[
  r_1 = \frac{1 - \sqrt{1 - 4\alpha^2}}{2\alpha}, \quad r_2 = \frac{1 + \sqrt{1 - 4\alpha^2}}{2\alpha} \tag{15.11}
\]
Since \( \alpha \in (\frac{1}{2}, 1) \), the roots are complex conjugates so either \( f(x) = 0 \) for all \( x \in \mathbb{N} \) or \( f(x) < 0 \) for infinitely many \( x \in \mathbb{N}_+ \), in either case a contradiction. In fact, we can give the solution explicitly. Since \( 1 < 1/\alpha < 2 \), we can write \( 1/\alpha = 2 \cos \theta \) for some \( \theta \in (0, \pi/3) \). By Theorem 15.2,
\[
f(x) = f(0) P_x \left( \frac{1}{\alpha} \right) = f(0) \frac{\sin[(x+1)\theta]}{\sin \theta}, \quad x \in \mathbb{N}
\]
Next, suppose that \( X \) has the geometric distribution on \( \mathbb{N} \) with success parameter \( p \in (0, 1) \). So, the probability density function \( f \) is given by \( f(x) = p(1-p)^x \) for \( x \in \mathbb{N} \). Of course, \( X \) has an exponential distribution for the positive semigroup \((\mathbb{N}, +)\). Then \( F(0) = f(1) = p(1-p) \) and
\[
  F(x) = f(x-1) + f(x+1) = p(1-p)^{x-1} + p(1-p)^{x+1} = p(1-p)^{x-1}[1 + (1-p)^2], \quad x \in \mathbb{N}_+
\]
So the right rate function \( r \) of \( X \) for \((\mathbb{N}, \rightarrow)\) satisfies \( r(0) = f(0)/F(0) = 1/(1-p) \) and
\[
r(x) = \frac{f(x)}{F(x)} = \frac{1 - p}{1 + (1-p)^2}, \quad x \in \mathbb{N}_+
\]
Hence \( X \) has constant rate \((1-p)/[1 + (1-p)^2] \) on \( \mathbb{N}_+ \). Conversely, suppose that \( X \) has constant rate \( \alpha \) on \( \mathbb{N}_+ \), so that the probability density function \( f \) satisfies (15.10) (but not necessarily (15.9)). The only possible solution with \( f(x) \to 0 \) as \( x \to \infty \) is when \( \alpha < \frac{1}{2} \) so that there is a root \( r < 1 \) of the characteristic equation. The solution has the form \( f(x) = cr^x \). The requirement that \( f \) be a density function then implies that \( c = 1 - r \), so \( X \) has the geometric distribution with parameter \( p = 1 - r \).

Consider again the general case where \( X \) is a random variable with probability density function \( f \) on \( \mathbb{N} \) and with right probability function \( F \) for \((\mathbb{N}, \rightarrow)\). As usual, let \( Y = (Y_1, Y_2, \ldots) \) denote the random walk on \((\mathbb{N}, \rightarrow)\) associated with the distribution of \( X \), so that \( Y_1 \) has density function \( f \) and \( Y_n \to Y_{n+1} \) for each \( n \in \mathbb{N} \). The transition matrix \( P \) is given by \( P(0,1) = 1 \) and
\[
\begin{align*}
P(x, x-1) &= \frac{f(x-1)}{f(x-1) + f(x+1)}, \quad x \in \mathbb{N}_+ \\
P(x, x+1) &= \frac{f(x+1)}{f(x-1) + f(x+1)}, \quad x \in \mathbb{N}_+
\end{align*}
\]
Hence $Y$ is a birth-death chain on $\mathbb{N}$ with reflecting boundary point $0$. From Note 15.7, we know that $Y$ is positive recurrent and reversible, with invariant function $\varphi = fF$, so that $\varphi(0) = f(0)f(1)$ and

$$\varphi(x) = f(x)[f(x-1) + f(x+1)], \quad x \in \mathbb{N}_+$$

Of course, the general theory of birth-death chains would lead to the same conclusions. In the special case that $X$ has the geometric distribution with success parameter $p \in (0,1)$, so that $f(x) = p(1-p)^x$ for $x \in \mathbb{N}$, the invariant density function $\varphi$ is given by $\varphi(0) = p^2(1-p)$ and

$$\varphi(x) = p^2(1-p)2^{x-1}(2-p), \quad x \in \mathbb{N}_+$$

The geometric distribution is almost the most random way to put points in the infinite path $(\mathbb{N}, \rightarrow)$.

### 15.3 Finite Paths

For $m \in \mathbb{N}_+$, let $\mathbb{N}_m = \{0, 1, \ldots, m\}$, and define the relation $\rightarrow$ on $\mathbb{N}_m$ by $x \rightarrow x+1$ for $x \in \{0, 1, \ldots, m-1\}$ and $x \rightarrow x-1$ for $x \in \{1, 2, \ldots, m\}$. So the graph $(\mathbb{N}_m, \rightarrow)$ is simply the (undirected) path $(0, 1, \ldots, m)$ (with no loops). All of our objects of study depend on $m$, but we will suppress this from the notation unless necessary.

**Theorem 15.8.** The left walk function $\gamma_n$ of order $n \in \mathbb{N}$ for $(\mathbb{N}_m, \rightarrow)$ is given as follows: $\gamma_0(x) = 1$ for $x \in \mathbb{N}_m$ and then recursively,

$$
\begin{align*}
\gamma_{n+1}(0) &= \gamma_n(1) \\
\gamma_{n+1}(x) &= \gamma_n(x-1) + \gamma_n(x+1), \quad x \in \{1, 2, \ldots, m-1\} \\
\gamma_{n+1}(m) &= \gamma_n(m-1)
\end{align*}
$$

Also, $\gamma_n$ is symmetric: $\gamma_n(x) = \gamma_n(m-x)$ for $n \in \mathbb{N}$ and $x \in S$.

**Proof.** By definition, $\gamma_0(x) = 1$ for $x \in \mathbb{N}_m$ and $\gamma_{n+1} = \gamma_nK$ where as usual, $K$ is the adjacency matrix of $(\mathbb{N}_m, \rightarrow)$. In this case, for $u : \mathbb{N}_m \rightarrow \mathbb{R}$,

$$
\begin{align*}
uK(0) &= u(1) \\
\nuK(x) &= u(x-1) + u(x+1), \quad x \in \{1, 2, \ldots, m-1\} \\
\nuK(m) &= u(m-1)
\end{align*}
$$

So the result follows.

The coefficients $\gamma_n(x)$ for $m \in \mathbb{N}_+$, $n \in \mathbb{N}$, and $x \in \mathbb{N}_m$ are interesting, and are objects of study in graph theory. By definition, $\gamma_n(x)$ is the number of walks of length $n$ terminating in $x \in \mathbb{N}_m$ in the graph $(\mathbb{N}_m, \rightarrow)$. By symmetry, $\gamma_n(x)$ is also the number of walks of length $n$ starting at $x$. For most values of $m$ and $x$, it seems difficult to give a closed-form expression for $\gamma_n(x)$ for $n \in \mathbb{N}$ and $x \in \mathbb{N}_m$. The left generating function $\Gamma$ for $(\mathbb{N}_m, \rightarrow)$, which encodes the same information, is easier to study.

**Theorem 15.9.** For $x \in \mathbb{N}_m$, $\Gamma(x, t)$ is defined for $|t| < \frac{1}{2}$ and satisfies

$$
\begin{align*}
\Gamma(0,t) &= 1 + t\Gamma(1,t) \\
\Gamma(x,t) &= 1 + t\Gamma(x-1,t) + t\Gamma(x+1,t), \quad x \in \{1, 2, \ldots, m-1\}
\end{align*}
$$

Also, $x \rightarrow \Gamma(x, t)$ is symmetric: $\Gamma(x, t) = \Gamma(m-x, t)$ for $x \in \mathbb{N}_m$.

**Proof.** Recall that $\gamma_n(x)$ can be interpreted as the number of walks of length $n$ starting from $x \in S$. Since there are at most 2 neighbors of each state, $\gamma_n(x) \leq 2^n$. So the results follow from the definition of the left generating function

$$
\Gamma(x,t) = \sum_{n=0}^{\infty} \gamma_n(x)t^n, \quad x \in S
$$

and the result in Theorem 15.8.
So $\Gamma$ satisfies the relation given in Theorem 15.4 for the infinite walk, but only for $x \in \mathbb{N}_m$. Our next goal is to identify the (right) eigenvalues and eigenfunctions of $(\mathbb{N}_m, \rightarrow)$, and once again, the polynomials in Definition 15.1 will play a critical role.

**Theorem 15.10.** The characteristic polynomial of $(\mathbb{N}_m, \rightarrow)$ is $P_{m+1}$. If $\beta \in \mathbb{R}$ is an eigenvalue, then an eigenfunction $g$ is given by $g(x) = P_x(\beta)$ for $x \in \mathbb{N}_m$.

**Proof.** Let $K$ denote the adjacency matrix of $(\mathbb{N}_m, \rightarrow)$ as usual. Then $\beta \in \mathbb{R}$ is an eigenvalue and $g : \mathbb{N}_m \rightarrow \mathbb{R}$ a corresponding eigenfunction if $g$ is nonzero and $Kg = \beta g$, or equivalently,

$$g(1) = \beta g(0) \quad (15.12)$$
$$g(x + 1) = \beta g(x) - g(x - 1), \quad x \in \{1, 2, \ldots, m\} \quad (15.13)$$
$$g(m - 1) = \beta g(m) \quad (15.14)$$

Equivalently, we can consider functions $g : \mathbb{N} \rightarrow \mathbb{R}$ satisfying $(15.12)$ and $(15.13)$ but with the boundary condition $g(m + 1) = 0$ which then gives $(15.14)$. But by Theorem 15.5, the solution of $(15.12)$ and $(15.13)$ is given by $g(x) = g(0)P_x(\beta)$ for $x \in \mathbb{N}$, and the requirement that $g(m + 1) = 0$ means that $\beta$ is a root of $P_{m+1}$. \hfill \Box

**Corollary 15.1.** The eigenvalues of $(\mathbb{N}_m, \rightarrow)$ are

$$\beta_k = 2 \cos \left( \frac{k}{m + 2} \pi \right), \quad k \in \{1, 2, \ldots, m + 1\}$$

The eigenvalues are simple and for $k \in \{1, 2, \ldots, m + 1\}$, an eigenfunction $g_k$ corresponding to $\beta_k$ is given by

$$g_k(x) = \frac{\sin \left( \frac{(x+1)k}{m+2} \pi \right)}{\sin \left( \frac{k}{m+2} \pi \right)}, \quad x \in \mathbb{N}_m$$

**Proof.** These results follow from Theorem 15.1 and Theorem 15.2. \hfill \Box

Suppose now that $X$ is a random variable with values in $\mathbb{N}_m$ and with density function $f$. The right probability function $F$ of $f$ for $(\mathbb{N}_m, \rightarrow)$ is given by

$$F(0) = f(1)$$
$$F(x) = f(x - 1) + f(x + 1), \quad x \in \{1, 2, \ldots, m - 1\}$$
$$F(m) = f(m - 1)$$

As usual, our support assumption that $F(x) > 0$ for $x \in \mathbb{N}_m$ is in effect. The right rate function $r$ of $f$ for $(\mathbb{N}_m, \rightarrow)$ is given by

$$r(0) = \frac{f(0)}{f(1)}$$
$$r(x) = \frac{f(x)}{f(x - 1) + f(x + 1)}, \quad x \in \{1, 2, \ldots, m - 1\}$$
$$r(m) = \frac{f(m)}{f(m - 1)}$$

The right probability function determines the distribution for some values of $m \in \mathbb{N}_+$, but not others.

**Theorem 15.11.** Suppose that $F$ is the right probability function for $(\mathbb{N}_m, \rightarrow)$ of a distribution on $\mathbb{N}_m$. Then $F$ determines the distribution except when $m = 2 \mod 4$.

**Proof.** From Corollary 15.1, note that 0 is an eigenvalue if and only if $m$ is even. In this case, an eigenfunction $g$ is given by

$$g(2x) = (-1)^x, \quad x \in \{0, 1, \ldots, m/2\}$$
$$g(2x + 1) = 0, \quad x \in \{0, 1, \ldots, m/2 - 1\}$$

If $m$ is a multiple of 4 then $\sum_{x=0}^{m} g(x) = 1$. If $m = 2 \mod 4$ then $\sum_{x=0}^{m} g(x) = 0$. So the result follows from Corollary 4.1. \hfill \Box
Exercise 15.1. Consider the case $m = 2$ so that $(N_2, \rightarrow)$ is the undirected path $(0, 1, 2)$. Find two probability density functions that have the same upper probability function.

The equations for the density $f$ to have constant rate $\alpha \in (0, \infty)$ for $(N_m, \rightarrow)$ are

\begin{align*}
f(0) &= \alpha f(1) \quad \text{(15.15)} \\
f(x) &= \alpha [f(x - 1) + f(x + 1)], \quad x \in \{1, 2, \ldots, m - 1\} \quad \text{(15.16)} \\
f(m) &= \alpha f(m - 1) \quad \text{(15.17)}
\end{align*}

and of course we also need $f(x) \geq 0$ for $x \in S$ and $\sum_{x=0}^{m} f(x) = 1$. Here is our main result:

**Theorem 15.12.** For each $m \in \mathbb{N}_+$, there exists a unique constant rate distribution for $(N_m, \rightarrow)$. The rate is

$$\alpha = \frac{1}{2 \cos \left( \frac{\pi}{m+2} \right)}$$

and the density function $f$ is given by

$$f(x) = \frac{\sin \left( \frac{x}{m+2} \right)}{1 + \cos \left( \frac{\pi}{m+2} \right)} \sin \left( \frac{x + 1}{m + 2} \pi \right), \quad x \in S$$

**Proof.** The eigenvalues of the adjacency matrix $K$ are given in Corollary 15.10. The largest eigenvalue is

$$2 \cos \left( \frac{\pi}{m+2} \right)$$

and a corresponding eigenfunction $g$ is given by

$$g(x) = \sin \left( \frac{x + 1}{m + 2} \pi \right), \quad x \in S$$

So the result follows from the general Theorem 4.10. The probability density function $f$ is the normalized eigenfunction.

Graphs of the constant rate density function are given in Figure 15.1 for $m = 10$ and $m = 100$.

![Figure 15.1: The graphs of $f$ with $m = 10$ and $m = 100$](image)

Exercise 15.2. Consider the case $m = 2$ so that $(N_2, \rightarrow)$ is the undirected path $(0, 1, 2)$.

(a) Find the left walk functions.

(b) Find the left generating function.

(c) Find the rate constant and the probability density function of the constant rate distribution.

(d) For $n \in \{2, 3, \ldots\}$, find the order $n$ probability density function corresponding to the constant rate distribution.
Exercise 15.3. Consider the case \( m = 3 \) so that \((\mathbb{N}_2, \rightarrow)\) is the undirected path \((0, 1, 2, 3)\).

(a) Find the left walk functions.

(b) Find the left generating function.

(c) Find the rate constant and the probability density function of the constant rate distribution.

(d) For \( n \in \{2, 3, \ldots\} \), find the order \( n \) probability density function corresponding to the constant rate distribution.

Exercise 15.4. Consider the case \( m = 4 \) so that \((\mathbb{N}_2, \rightarrow)\) is the undirected path \((0, 1, 2, 3, 4)\).

(a) Find the left walk functions.

(b) Find the left generating function.

(c) Find the rate constant and the probability density function of the constant rate distribution.

(d) For \( n \in \{2, 3, \ldots\} \), find the order \( n \) probability density function corresponding to the constant rate distribution.

Note that \( f \) is symmetric so that \( f(x) = f(m - x) \) for \( x \in \mathbb{N}_m \). Also, \( f(0) = 1 - \frac{1}{2^m} \) so that \( \alpha \in \left(\frac{1}{2}, 1\right) \).

(This also follows from the same proof as in Theorem 15.7). Finally, note that \( \alpha \to \frac{1}{2} \) as \( m \to \infty \). Next we give the asymptotic behavior of the constant rate distributions.

Theorem 15.13. Suppose that \( X_m \) has the constant rate distribution on \((\mathbb{N}_m, \rightarrow)\) for each \( m \in \mathbb{N}_+ \). Then as \( m \to \infty \), the distribution of \( \frac{X_m}{m} \) converges (in the usual sense) to the Gilbert’s sine distribution on \([0, 1]\), with density function

\[
x \mapsto \frac{\pi}{2} \sin(\pi x), \quad x \in [0, 1]
\]

Proof. Let \( x \in [0, 1] \). Then

\[
P(X_m/m \leq x) = \sum_{k=0}^{m} \mathbf{1}\left(\frac{k}{m} \leq x\right) f(k)
\]

\[
= \frac{\sin\left(\frac{\pi x}{m+2}\right)}{1 + \cos\left(\frac{\pi x}{m+2}\right)} \sum_{k=0}^{m} \mathbf{1}\left(\frac{k}{m} \leq x\right) \sin\left(\frac{k+1}{m+2}\pi\right)
\]

\[
= \left[ m \sin\left(\frac{\pi x}{m+2}\right) \right] \left[ \frac{1}{m} \sum_{k=0}^{m} \mathbf{1}\left(\frac{k}{m} \leq x\right) \sin\left(\frac{k+1}{m+2}\pi\right) \right]
\]

The factor in the first set of square brackets converges to \( \pi/2 \) as \( m \to \infty \) by simple calculus. The factor in the second set of square brackets is a Riemann sum for

\[
\int_{0}^{x} \sin(\pi u)du = \frac{1}{\pi} \left[ 1 - \cos(\pi x) \right]
\]

Hence

\[
P(X_m/m \leq x) \to \frac{1}{2} \left[ 1 - \cos(\pi x) \right] \text{ as } m \to \infty
\]

As a function of \( x \in [0, 1] \), the right side is the right probability function (in the ordinary sense) of Gilbert’s sine distribution.

The graph of Gilbert’s sine density function is given in Figure 15.2
15.4 Grid Graphs

For \( k \in \mathbb{N}_+ \), let \( \mathbb{N}_k = \{0, 1, \ldots, k\} \) as before.

**Definition 15.2.** For \( m, n \in \mathbb{N}_+ \), the \( m \times n \) grid graph \((\mathbb{N}_m \times \mathbb{N}_n, \rightarrow)\) is the Cartesian (graph) product of the path graphs \((\mathbb{N}_m, \rightarrow)\) and \((\mathbb{N}_n, \rightarrow)\). That is,

\[
\begin{align*}
(x, y) \rightarrow (x+1, y), & \quad x \in \{0, 1, \ldots, m-1\}, y \in \{0, 1, \ldots, n\} \\
(x, y) \rightarrow (x-1, y), & \quad x \in \{1, 2, \ldots, m\}, y \in \{0, 1, \ldots, n\} \\
(x, y) \rightarrow (x, y+1), & \quad x \in \{0, 1, \ldots, m\}, y \in \{0, 1, \ldots, n-1\} \\
(x, y) \rightarrow (x, y-1), & \quad x \in \{0, 1, \ldots, m\}, y \in \{1, 2, \ldots, n\}
\end{align*}
\]

For \( k \in \mathbb{N}_+ \), let \( \alpha_k \) denote the rate constant and \( f_k \) the density function of the constant rate distribution on the path \((\mathbb{N}_k, \rightarrow)\). That is,

\[
\alpha_k = \frac{1}{2 \cos \left( \frac{\pi}{k+2} \right)}
\]

\[
f_k(x) = \frac{\sin \left( \frac{\pi}{k+2} \right)}{1 + \cos \left( \frac{\pi}{k+2} \right)} \sin \left( \frac{x+1}{k+2} \pi \right), \quad x \in \mathbb{N}_k
\]

Here is the main result of this section:

**Theorem 15.14.** The \( m \times n \) grid graph \((\mathbb{N}_m \times \mathbb{N}_n, \rightarrow)\) has a unique constant rate distribution with rate

\[
\alpha_{m,n} = \frac{\alpha_m \alpha_n}{\alpha_m + \alpha_n} = \frac{1}{2 \cos \left( \frac{\pi}{m+2} \right) + 2 \cos \left( \frac{\pi}{n+2} \right)}
\]

and with density function \( f_{m,n} \) given by

\[
f_{m,n}(x, y) = f_m(x)f_n(y)
\]

\[
= \frac{\sin \left( \frac{\pi}{m+2} \right) \sin \left( \frac{\pi}{n+2} \right)}{\left[ 1 + \cos \left( \frac{\pi}{m+2} \right) \right] \left[ 1 + \cos \left( \frac{\pi}{n+2} \right) \right]} \sin \left( \frac{x+1}{m+2} \pi \right) \sin \left( \frac{y+1}{n+2} \pi \right), \quad (x, y) \in \mathbb{N}_m \times \mathbb{N}_n
\]

**Proof.** This follows immediately from the general result on Cartesian products of graphs. \( \blacksquare \)
Chapter 16

The Subset Semigroup

16.1 Preliminaries

Let $S$ denote the set of all finite subsets of $\mathbb{N}_+$. Our primary interest in this chapter is the discrete partial order graph $(S, \subseteq)$. Interestingly, this graph is associated with a positive semigroup, so we can discuss the exponential and memoryless properties, in addition to the constant rate property. We also consider the strict partial order graph $(S, \subset)$, the covering graph $(S, \uparrow)$ associated with $(S, \subseteq)$, and finally, the reflexive completion $(S, \uparrow\uparrow)$ associated with $(S, \uparrow)$. First we give some preliminary results for the various graphs. Recall the permutation (falling power) formula: for $m \in \mathbb{R}$ and $k \in \mathbb{N}$,

$$m^{(k)} = m(m-1) \cdots (m-k+1)$$

**Proposition 16.1.** For $n \in \mathbb{N}$, the left walk function $w_n$ of $(S, \subseteq)$ is given by

$$w_n(x) = (n+1)^\#(x), \quad x \in S$$

**Proof.** Here is a combinatorial proof: A walk in $(S, \subseteq)$ of length $n$ terminating in $x$ is a sequence $(x_1, x_2, \ldots, x_{n+1})$ of sets in $S$ with $x_i \subseteq x_{i+1}$ for $i \in \{1, 2, \ldots, n\}$ and $x_{n+1} = x$. The following algorithm generates all such sequences once and only once: For each $k \in x$ either $k$ is not in any of the sets in the sequence $(x_1, x_2, \ldots, x_n)$ or $k$ occurs in the sequence for the first time in set $x_i$ for some $i \in \{1, 2, \ldots, n\}$. Thus there are $n+1$ choices for each element of $x$.

Here is a proof by induction: The result is trivially true when $n = 0$, since $w_0(x) = 1$ for $x \in S$. Thus, assume the result holds for a given $n$. Using the binomial theorem,

$$w_{n+1}(x) = \sum_{t \subseteq x} w_n(t) = \sum_{k=0}^{\#(x)} \sum_{t \subseteq x, \#(t)=k} (n+1)^k$$

$$= \sum_{k=0}^{\#(x)} \binom{\#(x)}{k} (n+1)^k = (n+2)^{\#(x)}, \quad x \in S$$

When $n = 1$, Proposition 16.1 gives the usual formula for the number of subsets of $x$, namely $w_1(x) = 2^{\#(x)}$.

**Corollary 16.1.** The left walk functions of order $n \in \mathbb{N}$ for the other graphs are as follow:

(a) For the graph $(S, \subset)$,

$$w_n(x) = \sum_{k=0}^{\#(x)} (-1)^{n-k} \binom{n}{k} (k+1)^{\#(x)}, \quad x \in S$$

(b) For the graph $(S, \uparrow)$,

$$w_n(x) = \#(x)^{(n)}, \quad x \in S$$
(c) For the graph \((S, \uparrow)\),

\[ w_n(x) = \sum_{k=0}^{n} \binom{n}{k} \#(x)^{(k)}, \quad x \in S \]

**Proof.** The proofs are straightforward.

(a) This follows from Proposition 16.1 and the results for reflexive completion.

(b) Let \(x \in S\) and \(n \in \mathbb{N}\) with \(n \leq \#(x)\). To form a walk of length \(n\) in \((S, \uparrow)\) terminating in \(x\), select a permutation \((k_1, k_2, \ldots, k_n)\) of size \(n\) from \(x\) and let \(y = x \setminus \{k_1, k_2, \ldots, k_n\}\). The walk is then

\[ (y, y \cup \{k_1\}, y \cup \{k_1, k_2\}, \ldots, y \cup \{k_1, k_2, \ldots, k_n\} = x) \]

The number of ways of selecting a permutation of size \(n\) from \(x\) is \(\#(x)^{(n)}\).

(c) This follows from (b) and results for reflexive completion.

Next we give the left generating functions. For the following theorem, recall that \(\text{Li}\) denotes the polylogarithm function and we will let \(\Gamma\) denote the incomplete gamma function.

**Proposition 16.2.** The left generating function \(W\) of \((S, \subseteq)\) is given by

\[ W(x, t) = \frac{1}{t} \text{Li}\left[-\#(x), t\right], \quad x \in S, \quad t \in (-1, 1) \]

**Proof.**

\[ W(x, t) = \sum_{n=0}^{\infty} w_n(x) t^n = \sum_{n=0}^{\infty} (n + 1)^\#(x) t^n = \frac{1}{t} \text{Li}\left[-\#(x), t\right], \quad t \in (-1, 1) \]

where it is understood that \(W(x, 0) = 1\).

**Corollary 16.2.** The left generating functions for the other graphs are as follows:

(a) For the graph \((S, \subset)\),

\[ W(x, t) = \frac{1}{t} \text{Li}\left[-\#(x), \frac{t}{1+t}\right], \quad x \in S, \quad t \in \mathbb{R} \]

(b) For the graph \((S, \uparrow)\),

\[ W(x, t) = e^{1/t} t^\#(x) \Gamma[1 + \#(x), 1/t] \]

(c) For the graph \((S, \uparrow)\)

\[ W(x, t) = \frac{1}{1 - t} e^{(1-t)/t} \left(\frac{t}{1-t}\right)^\#(x) \Gamma\left[1 + \#(x), \frac{1-t}{t}\right] \]

**Proof.** The proofs are straightforward.

(a) This follows from standard results on reflexive completion.

(b) For \(x \in S\),

\[ W(x, t) = \sum_{n=0}^{\#(x)} [\#(x)]^{(n)} t^n = e^{1/t} t^\#(x) \Gamma(1 + m, 1/t), \quad t \in \mathbb{R} \]

(c) This follows from standard results on reflexive completion.
Note that the left generating function for \((S, η)\) is a simple polynomial: \(W(x, t) = φ_{\#(x)}(t)\) where
\[
φ_m(t) = 1 + mt + m(m-1)t^2 + \cdots + mt^m, \quad m \in \mathbb{N}, \ t \in \mathbb{R}
\]
The polynomials \(φ_m\) for \(m \in \mathbb{N}\) can be generated recursively by
\[
φ_{m+1}(t) = 1 + (m+1)tφ_m(t), \quad t \in \mathbb{R}
\]
along with the initial condition \(φ_0(t) = 1\) for all \(t \in \mathbb{R}\).

**Proposition 16.3.** the Möbius function \(M\) for \((S, ⊆)\) is given by
\[
M(x, y) = (-1)^{\#(y) - \#(x)}, \quad (x, y) ∈ S^2, \ x ≤ y
\]

**Proof.** This is a well-known result, but we give a proof for completion, based on induction on \(\#(y) - \#(x)\). First, \(M(x, x) = 1\) by definition. Suppose that the result holds for \(x ≤ y\) when \(\#(y) = \#(x) + n\). Suppose that \(x ≤ y\) with \(\#(y) ≤ \#(x) + n + 1\). Then
\[
M(x, y) = - \sum_{t ∈ [x, y)} M(x, t) = - \sum_{k=0}^{n} \sum_{t ∈ [x, y)} \{M(x, t) : t ∈ [x, y), \#(t) = \#(x) + k\}
\]
\[
= - \sum_{k=0}^{n} \sum_{t ∈ [x, y)} \{(-1)^k : t ∈ [x, y), \#(t) = \#(x) + k\}
\]
\[
= - \sum_{k=0}^{n} \binom{n+1}{k} (-1)^k
\]
\[
= - \sum_{k=0}^{n+1} \binom{n+1}{k} (-1)^k + (-1)^{n+1} = 0 + (-1)^{n+1}
\]

We can now verify the Möbius inversion formula in this special case, with the left walk function \(w_n\) of order \(n ∈ \mathbb{N}\) for \((S, ⊆)\). Recall that \(w_n(x) = (n + 1)^{\#(x)}\) for \(x ∈ S\):
\[
\sum_{t ≤ x} w_{n+1}(t)M(t, x) = \sum_{t ≤ x} (n + 2)^{\#(x)}(-1)^{\#(x) - \#(t)}
\]
\[
= \sum_{k=0}^{\#(x)} \sum_{t ≤ x} \{(n + 2)^{\#(x)}(-1)^{\#(x) - \#(t)} : t ≤ x, \#(t) = k\}
\]
\[
= \sum_{k=0}^{\#(x)} \binom{\#(x)}{k} (n + 2)^k (-1)^{\#(x) - k}
\]
\[
= (n + 1)^{\#(x)} = w_n(x), \quad x ∈ S
\]

### 16.2 The Positive Semigroup

Constructing a semigroup \((S, \cdot)\) that corresponds to the graph \((S, ⊆)\) requires a bit of work. First, we identify a nonempty subset \(x \in \mathbb{N}_+\) with the function given by
\[
x(i) = \text{ith smallest element of } x
\]
with domain \(D(x) = \{1, 2, \ldots, \#(x)\}\) if \(x\) is finite and with domain \(\mathbb{N}_+\) if \(x\) is infinite. We will sometimes refer to \(x(i)\) as the element of rank \(i\) in \(x\). If \(x\) is nonempty, \(\min(x)\) denotes the minimum of \(x\); by convention we take \(\min(\emptyset) = \infty\). If \(x\) is nonempty and finite, \(\max(x)\) denotes the maximum of \(x\); by convention we take \(\max(\emptyset) = 0\) and \(\max(x) = \infty\) if \(x\) is infinite. Note that \(\#(x) ≤ \max(x)\) for every \(x\).
Definition 16.1. If \( x \) and \( y \) are nonempty subsets of \( \mathbb{N}_+ \) with \( \max(y) \leq \#(x) \), let \( x \circ y \) denote the subset of \( \mathbb{N}_+ \) whose function is the composition of \( x \) and \( y \):

\[
(x \circ y)(i) = x[y(i)], \quad i \in D(y)
\]

We also define \( x \circ \emptyset = \emptyset \) for any \( x \subseteq \mathbb{N}_+ \).

Note that \( x \circ y \) is always defined when \( x \) is infinite. The results in the following two propositions are simple:

**Proposition 16.4.** Suppose that \( x \) and \( y \) are subsets of \( \mathbb{N}_+ \) with \( \max(y) \leq \#(x) \). Then

(a) \( x \circ y \subseteq x \)

(b) \( \#(x \circ y) = \#(y) \)

(c) If \( y \) is nonempty and finite then \( \max(x \circ y) = x[\max(y)] \)

(d) If \( x \) is infinite then \( (x \circ y)^c = x^c \cup (x \circ y)^c \)

*Proof.* Parts (a), (b), and (c) are clear since \( x \circ y \) consists of the elements of \( x \) that are indexed by the elements of \( y \). Suppose that \( x \subseteq \mathbb{N}_+ \) is infinite, so that \( x \circ y \) is defined for every \( y \subseteq \mathbb{N}_+ \). For part (d), note that by definition an element \( j \notin x \circ y \) if and only if either \( j \notin x \) or \( j \in x \), but \( j \) is not indexed by an element of \( y \) and hence is indexed by an element of \( y^c \).

**Proposition 16.5.** Suppose that \( x, y, \) and \( z \) are subsets of \( \mathbb{N}_+ \). Assuming that the operations are defined,

(a) \( x \circ (y \circ z) = (x \circ y) \circ z \)

(b) \( x \circ (y \cup z) = (x \circ y) \cup (x \circ z) \)

(c) \( x \circ (y \cap z) = (x \circ y) \cap (x \circ z) \)

(d) If \( x \circ y = x \circ z \) then \( y = z \).

*Proof.* The proofs are straightforward.

(a) Note that \( x \circ (y \circ z) \) and \( (x \circ y) \circ z \) are both defined if \( \max(y) \leq \#(x) \) and \( \max(z) \leq \#(y) \). In this case,

\[
[x \circ (y \circ z)][i] = [(x \circ y) \circ z][i] = x[y[z(i)]], \quad i \in D(z)
\]

In general, of course, the composition of functions is associative when the compositions make sense.

(b) Note that \( x \circ (y \cup z) \) and \( (x \circ y) \cup (x \circ z) \) are both defined if \( \max(y) \leq \#(x) \) and \( \max(z) \leq \#(x) \). In this case, the elements of \( x \) indexed by \( y \cup z \) is the union of the elements of \( x \) indexed by \( y \) with the elements of \( x \) indexed by \( z \).

(c) Again, \( x \circ (y \cap z) \) and \( (x \circ y) \cap (x \circ z) \) are both defined if \( \max(y) \leq \#(x) \) and \( \max(z) \leq \#(x) \). In this case, the elements of \( x \) indexed by \( y \cap z \) is the intersection of the elements of \( x \) indexed by \( y \) with the elements of \( x \) indexed by \( z \).

(d) If \( x \circ y = x \circ z \) then \( \#(y) = \#(z) \), so let \( d \) denote the common value. Also, \( \max(y) \leq \#(x) \) and \( \max(z) \leq \#(x) \). By definition, \( x[y(i)] = x[z(i)] \) for \( i \in \{1, 2, \ldots, d\} \). Hence \( y = z \).

Part (a) is the associate law for composition; part (b) is the left distributive law for composition over union; and part (c) is the left distributive law for composition over intersection. Finally, part (d) is the left cancellation law. Note that the right distributive laws cannot possibly hold; \( (x \cup y) \circ z \) and \( (x \circ z) \cup (y \circ z) \) do not even have the same cardinality in general, and neither do \( (x \cap y) \circ z \) and \( (x \circ z) \cap (y \circ z) \). Similarly, the right cancellation law does not hold: if \( x \circ z = y \circ z \), we cannot even conclude that \( \#(x) = \#(y) \), let alone that \( x = y \). Note that \( \mathbb{N}_+ \) is a left identity: \( \mathbb{N}_+ \circ x = x \) for any \( x \subseteq \mathbb{N}_+ \).

**Exercise 16.1.** Let \( x = \{2, 5, 6, 8, 12, 13, 25\} \), \( y = \{1, 4, 7\} \), \( z = \{1, 3, 4, 6\} \). Find each of the following sets directly from the definition:
Exercise 16.2. Let $x$ denote the set of odd positive integers and $y$ the set of even positive integers. Find each of the following sets:

(a) $x \circ x$
(b) $x \circ (y \cup z)$, $(x \circ y) \cup (x \circ z)$
(c) $x \circ (y \cap z)$, $(x \circ y) \cap (x \circ z)$

We can now define semigroup.

Theorem 16.1. Define the binary operation $\cdot$ on $S$ by
\[ xy = x \cup (x^c \circ y) = x \cup \{i\text{th smallest element of } x^c : i \in y\} \]

Then $(S, \cdot)$ is a positive semigroup with associated graph $(S, \subseteq)$.

Proof. Note that the operation is well defined since $x^c$ is infinite for $x \in S$. Essentially, $xy$ is constructed by adding to $x$ those elements of $x^c$ that are indexed by $y$ (in a sense those elements form a copy of $y$ that is disjoint from $x$). The associative rule holds, and in fact for $x, y, z \in S$,
\[ x(yz) = (xy)z = x \cup (x^c \circ y) \cup (x^c \circ y^c \circ z) \]

The empty set is the identity: If $x \in S$ then
\[ x\emptyset = x \cup (x^c \circ \emptyset) = x \cup \emptyset = x \]
\[ \emptyset x = \emptyset \cup (\mathbb{N}_+ \circ x) = \emptyset \cup x = x \]

The left cancellation law holds: Suppose that $x, y, z \in S$ and that $xy = xz$. Then $x \cup (x^c \circ y) = x \cup (x^c \circ z)$ by definition and hence $x^c \circ y = x^c \circ z$ since the pairs of sets in each union are disjoint. But then $y = z$. There are no non-trivial inverses: If $x, y \in S$ and $xy = \emptyset$ then $x \cup (x^c \circ y) = \emptyset$. Hence we must have $x = \emptyset$ and therefore also $x^c \circ y = \mathbb{N}_+ \circ y = y = \emptyset$. Finally, the associated partial order is the subset order. Suppose first that $x, u, y \in S$ and that $xu = y$. Then $x \cup (x^c \circ u) = y$ so $x \subseteq y$. Conversely, suppose that $x, y \in S$ and that $x \subseteq y$. Let $u = \{i \in \mathbb{N}_+: x^c(i) \in y\}$. Then $u \in S$ and $x \cup (x^c \circ u) = y$ so $xu = y$.\[\]

With our usual notation, $S_+ = \{x \in S : x \neq \emptyset\}$, so that $(S_+, \cdot)$ is the corresponding strict positive semigroup. The irreducible elements of $(S, \cdot)$ are the singletons $\{i\}$ where $i \in \mathbb{N}_+$. Note also that
\[ \{i\} \{i\} = \{i, i + 1\} \]
\[ \{i + 1\} \{i\} = \{i, i + 1\} \]
\[ \{i\} \{i + 1\} = \{i, i + 2\} \]

Comparing (16.2) and (16.3) we see that the semigroup is not commutative, and comparing (16.1) and (16.2) we see that the right cancellation law does not hold. Thus, $(S, \cdot)$ just satisfies the minimal algebraic assumptions of a positive semigroup; in particular, $S$ cannot be embedded in a group. A couple of other facts of interest are
(a) If $i_1, i_2, \ldots, i_n \in \mathbb{N}_+$ with $i_1 < i_2 < \cdots < i_n$ then $\{i_n\} \{i_{n-1}\} \cdots \{i_1\} = \{i_1, i_2, \ldots, i_n\}$.
(b) If $i \in \mathbb{N}_+$ then $\{i\}^n = \{i, i + 1, \ldots, i + n - 1\}$.\[\]
Exercise 16.3. Let $x = \{2, 3, 10\}$ and $y = \{4, 7\}$. Find each of the following:

(a) $xy$
(b) $yx$
(c) $x^2$
(d) $y^2$

The other relations of interest can also be defined in terms of the operator $\cdot$. Let $x, y \in S$. Then $x \subset y$ if and only if $xz = y$ for some nonempty $z \in S$. Next, $x \uparrow y$ if and only if $y = x\{i\}$ for some $i \in \mathbb{N}_+$. Finally, $x \uparrow y$ if and only if $x = y$ or $y = x\{i\}$ for some $i \in \mathbb{N}_+$.

Proposition 16.6. If $\varphi$ is a homomorphism from $(S, \cdot)$ into $(\mathbb{R}, +)$, then $\varphi = c\#$ for some $c \in \mathbb{R}$. In particular, $\text{dim}(S, \cdot) = 1$

Proof. Suppose that $\varphi$ is a homomorphism from $(S, \cdot)$ into $(\mathbb{R}, +)$. Then $\varphi(\{i\}\{i\}) = \varphi(\{i+1\}\{i\})$ and hence $\varphi(\{i\}) + \varphi(\{i\}) = \varphi(\{i+1\}) + \varphi(\{i\})$. It follows that $\varphi(\{i+1\}) = \varphi(\{i\})$ for $i \in \mathbb{N}_+$ and so $\varphi(\{i\}) = \varphi(\{1\})$ for $i \in \mathbb{N}_+$. Let $c = \varphi(\{1\})$. If $i_1 < i_2 < \cdots < i_n$ are positive integers then

\[
\varphi(\{i_1, i_2, \ldots, i_n\}) = \varphi(\{i_n\}\{i_{n-1}\} \cdot \cdots \{i_1\}) = \varphi(\{i_n\}) + \varphi(\{i_{n-1}\}) + \cdots + \varphi(\{i_1\}) = cn
\]

So $\varphi(x) = c\#(x)$ for $x \in S$. In particular, if $\varphi(\{1\}) = 0$ then $\varphi(x) = 0$ for all $x \in S$. On the other hand, if $c \neq 0$ then $c\#$ is a non-trivial homomorphism.

Proposition 16.7. For $x, y \in S$,

(a) $\#(xy) = \#(x) + \#(y)$
(b) $\max(xy) = \begin{cases} \max(x), & \text{if } \max(y) \leq \max(x) - \#(x) \\ \max(y) + \#(x), & \text{if } \max(y) > \max(x) - \#(x) \end{cases}$
(c) $\min(xy) = \min\{\min(x), \min(y)\}$

Proof. Let $x, y \in S$.

(a) $\#(xy) = \#(x \cup (x^c \circ y)) = \#(x) + \#(x^c \circ y)$ since $x$ and $x^c \circ y$ are disjoint. But from Theorem 27, $\#(x^c \circ y) = \#(y)$.

(b) The result is trivial if $x$ or $y$ is the identity ($\emptyset$), so we will assume that $x$ and $y$ are nonempty. Note that, by definition

\[
\max(xy) = \max(x \cup (x^c \circ y)) = \max\{\max(x), \max(x^c \circ y)\}
\]

Let $i = \#(x)$ and $n = \max(x)$. Then $n \in x$ and the remaining $i - 1$ elements of $x$ are in $\{1, 2, \ldots, n-1\}$. Hence, $x^c$ contains $n-i$ elements of $\{1, 2, \ldots, n-1\}$, together with all of the elements of $\{n+1, n+2, \ldots\}$. If $\max(y) \leq n-i$, then $\max(x^c \circ y) = x^c[\max(y)] \leq n-1$ so $\max(xy) = n = \max(x)$. If $\max(y) > n-i$, then $\max(xy) = \max(x^c \circ y) = x^c[\max(y)]$—the element of rank $\max(y)$ in $x^c$. Given the structure of $x^c$ noted above, this element is $n + [\max(y) - (n-i)] = \max(y) + i$.

(c) Again, the result is trivial if $x$ or $y$ is empty, so consider the case where both are nonempty and let $i = \min(x)$ and $j = \min(y)$. The smallest element of $x^c \circ y$ is $x^c(j)$. If $i \leq j$ then $x^c(j) > i$. If $i > j$ then $x^c(j) = j$.

Since the cardinality function is a homomorphism, the partial order graph $(S, \subseteq)$ is uniform. That is, if $x \in S$ can be factored into singletons

\[
x = u_1 u_2 \cdots u_n
\]

where $\#(u_i) = 1$ for each $i$, then $n = \#(x)$. Moreover, we know the number of such factorings.
Proposition 16.8. If \( x \in S \) with \( \#(x) = n \) then there are \( n! \) factorings of \( x \) into irreducible elements (singletons).

**Proof.** The proof is by induction on \( n \). The result is trivially true when \( n = 1 \) since \( x \) itself would be a singleton. Suppose that the result holds for a given \( n \) and suppose that \( x \in S \) with \( \#(x) = n + 1 \). Note that in general, if \( y, z \in S \), \( i \in \mathbb{N}_+ \) and \( z = \{i\}y \) then \( i \in Z \). So an algorithm for constructing a factoring of \( x \) is as follows: Select \( i \in x \) and select a factoring of \( \{i\}^{-1}x \). The first step can be performed in \( n + 1 \) ways, and by the induction hypothesis, the second step can be performed in \( n! \) ways. Hence the number of factorings of \( x \) is \( (n+1)! = (n+1)! \).

We can also give an explicit algorithm for the factorings. Let \( x = \{i_1, i_2, \ldots, i_n\} \in S \). Let \((j_1, j_2, \ldots, j_n)\) be a permutation of \((i_1, i_2, \ldots, i_n)\). For each \( m \in \{1, 2, \ldots, n\} \), let \( k_m = \#\{j \in \{j_1, \ldots, j_{m-1}\} : j > k\} \). Then the factoring of \( x \) corresponding to the given permutation is \( \prod_{m=1}^{n}(j_m - k_m) \). Of course, there are \( n! \) permutations of the elements of \( x \).

**Exercise 16.4.** Suppose that \( i, j, k \in \mathbb{N}_+ \) with \( i < j < k \).

(a) Give the two factorings of \( \{i, j\} \) into singletons.
(b) Give the six factorings of \( \{i, j, k\} \) into singletons.

**Exercise 16.5.** Let \( x = \{2, 3, 5, 12, 17\} \in S \). Give the factoring of \( x \) into singletons corresponding to the following permutations.

(a) \((3, 5, 17, 2, 12)\)
(b) \((2, 5, 3, 17, 12)\)

16.3 Sub-semigroups

The semigroup \((S, \cdot)\) has an interesting structure. In particular, it can be partitioned into sub-semigroups where the difference between the maximum element and the cardinality is constant.

**Definition 16.2.** For \( k \in \mathbb{N} \), let

\[
S_k = \{x \in S : \max(x) - \#(x) = k\}
\]

Let \( S_{0,0} = \emptyset \), and for \( n \in \mathbb{N}_+ \) and \( k \in \mathbb{N} \), let

\[
S_{n,k} = \{x \in S : \#(x) = n, \max(x) = n + k\} = \{x \in S_k : \#(x) = n\}
\]

Note that \( \emptyset \in S_0 \) but \( \emptyset \notin S_k \) for \( k \in \mathbb{N}_+ \). In fact,

\[
S_0 = \emptyset, \{1\}, \{1,2\}, \ldots
\]

If \( k \in \mathbb{N}_+ \), then \( S_k = \bigcup_{n=1}^{\infty} S_{n,k} \) and the sets in the union are disjoint. Moreover, for \( k \in \mathbb{N} \) and \( n \in \mathbb{N}_+ \),

\[
\#(S_{n,k}) = \binom{n+k-1}{n-1} \quad (16.4)
\]

since \( x \in S_{n,k} \) must contain the element \( n + k \) and \( n - 1 \) elements from \( \{1, 2, \ldots, n+k-1\} \). Suppose that \( x \in S_{n,k} \) and \( y \in S_{m,j} \) where \( j, k \in \mathbb{N} \) and \( m, n \in \mathbb{N}_+ \). If \( m + j \leq k \) (so that \( j \leq k - m \)), then from Proposition 16.7, \( \max(xy) = \max(x) = n + k \) and \( \#(xy) = n + m \). Therefore

\[
\max(xy) - \#(xy) = (n + k) - (n + m) = k - m \quad (16.5)
\]

so \( xy \in S_{n+m,k-m} \). On the other hand, if \( m + j > k \) (so that \( j > k - m \)) then \( \max(xy) = \max(y) + \#(x) = m + j + n \) and as before, \( \#(xy) = n + m \). Therefore

\[
\max(xy) - \#(xy) = (m + j + n) - (n + m) = j \quad (16.6)
\]

so \( xy \in S_{n+m,j} \).
Theorem 16.2. \((S_k, \cdot)\) is a complete sub-semigroup of \((S, \cdot)\) for each \(k \in \mathbb{N}\).

Proof. We first need to show that \(xy \in S_k\) for \(x, y \in S_k\). The result is trivial if \(x = 0\) or \(y = 0\) (which is only possible if \(k = 0\)), so we will assume that \(x\) and \(y\) are nonempty. Then \(\max(y) > \max(x) - \#(x)\), since the left-hand side is \(k + \#(y)\) and the right-hand side is \(k\). By Proposition 16.7, \(\max(xy) = \max(y) + \#(x)\).

Hence

\[
\max(xy) - \#(xy) = [\max(y) + \#(x)] - [\#(x) + \#(y)] = \max(y) - \#(y) = k
\]

Therefore \(xy \in S_k\).

To show completeness, suppose that \(x, y \in S_k\) and \(x \subseteq y\) so that \(xu = y\) for some \(u \in S_n\). If \(\max(u) \leq k\) then from Proposition 16.7, \(\max(y) = \max(x)\) and \(\#(y) = \#(x) + \#(u)\), so \(\max(y) - \#(y) = k - \#(u) < k\), a contradiction. Thus, \(\max(u) > k\). From Proposition 16.7 again, \(\max(u) = \max(y) - \#(x)\) and \(\#(u) = \#(y) - \#(x)\), so \(\max(u) - \#(u) = \max(y) - \#(y) = k\). Thus, \(u \in S_k\). □

Note that \((S_0, \cdot)\) is a positive semigroup since it contains the identity 0. On the other hand, \((S_k, \cdot)\) is a strict positive semigroup for \(k \in \mathbb{N}_+\). As noted earlier,

\[
S_0 = \{\{1, 2, \ldots, m\} : m \in \mathbb{N}\}
\]

If \(y \in S\) with \(\#(y) = n \in \mathbb{N}\) and if \(m \in \mathbb{N}\), then

\[
\{1, 2, \ldots, m\} y = \{1, 2, \ldots, m\} \cup \{m + y(1), m + y(2), \ldots, m + y(n)\}
\]

In particular,

\[
\{1, 2, \ldots, m\}\{1, 2, \ldots, n\} = \{1, 2, \ldots, m + n\}, \quad m, n \in \mathbb{N}
\]

so \((S_0, \cdot)\) is isomorphic to \((\mathbb{N}, +)\) with isomorphism \(x \mapsto \#(x)\). Since \((S_k, \cdot)\) is a complete, strictly positive sub-semigroup of \((S, \cdot)\) for \(k \in \mathbb{N}_+\), the relation associated with \((S_k, \cdot)\) is the strict subset relation \(\subset\). That is, for \(x, y \in S_k\), \(x \subseteq y\) if and only if \(x = zy\) for some \(z \in S_k\). Finally note that the collection \(\{S_k : k \in \mathbb{N}\}\) is disjoint. To characterize the exponential distributions on \((S_k, \cdot)\) for \(k \in \mathbb{N}\), we must first characterize the irreducible elements. For \(k \in \mathbb{N}_+\), the irreducible elements are also the minimal elements.

Theorem 16.3. For \(k \in \mathbb{N}\), the set of irreducible elements of \((S_k, \cdot)\) is

\[
M_k = \{x \in S_k : x(i) \leq k \text{ for all } i < \#(x)\}
\]

There are \(2^k\) minimal elements.

Proof. First we show that if \(x \in S_k\) is not an irreducible element of \((S_k, \cdot)\) then \(x \notin M_k\). Thus, suppose that \(x = uv\) where \(u, v \in S_k\). Then \(\max(u) > k\) and \(\max(u) \in u \subseteq uv = x\). Moreover, \(\max(u) < \max(x)\), so the rank of \(\max(u)\) in \(x\) is less than \(\#(x) = \#(u) + \#(v)\). Therefore \(x \notin M_k\).

Next we show that if \(x \in S_k - M_k\) then \(x\) is not an irreducible element of \((S_k, \cdot)\). Thus, suppose that \(x \in S_k\) and \(x(i) > k\) for some \(i < \#(x)\). Construct \(u \in S\) as follows: \(x(i) \in u\) and \(u\) contains \(x(i) - k\) elements of \(x\) that are smaller than \(x(i)\). This can be done since \(x(i) - i \leq k\), and hence \(x(i) - k - 1 \leq i - 1\), and by definition, \(x\) contains \(i - 1\) elements smaller than \(x(i)\). Now note that \(\max(u) - \#(u) = x(i) - [x(i) - k] = k\) so \(u \in S_k\). By construction, \(u \subseteq x\), so there exists \(v \in S\) such that \(uv = x\). Recall that \(v\) is the set of ranks of the elements of \(x - u\) in \(u^c\). But \(u^c\) contains \(k\) elements less than \(x(i)\) together with the elements \(x(i) + 1, x(i) + 2, \ldots\). The largest element of \(x - u\) is \(\max(x) = \#(x) + k\) which has rank greater than \(k\) in \(u^c\). Therefore \(\max(v) > k = \max(u) - \#(u)\) so by Theorem 16.7, \(\max(x) = \max(v) + \#(u)\). Therefore

\[
\max(v) - \#(v) = [\max(x) - \#(u)] - [\#(x) - \#(u)] = \max(x) - \#(x) = k
\]

so \(v \in S_k\). Therefore \(x\) is not an irreducible element of \(S_k\).

Next, note that if \(x \in S_k\) and \(\#(x) \geq k + 2\), then \(x \notin M_k\), since one of the \(k + 1\) elements of \(x\) of rank less than \(\#(x)\) must be at least \(k + 1\). For \(n \leq k + 1\), the number of elements \(x \in M_k\) with \(\#(x) = n\) is \(\binom{k}{n-1}\), since \(x\) must contain \(n + k\) and \(n - 1\) elements in \(\{1, 2, \ldots, k\}\). Hence

\[
\#(M_k) = \sum_{n=1}^{k+1} \binom{k}{n} = 2^k
\]
Exercise 16.6. Explicitly give the set of irreducible elements of \((S_k, \cdot)\) for \(k \in \{0, 1, 2, 3\}\).

Example 16.1. For \(k \in \mathbb{N}_+\), the number of elements in a factoring of an element in \(S_k\) into irreducible elements is not necessarily unique. For example, \(\{k+1\}, \{k, k+2\} \in M_k\) and

\[
\{k, k+2\} \{k, k+2\} = \{k+1\} \{k+1\} \{k, k+2\} = \{k, k+1, k+2, k+4\}
\]

Thus, \((S_k, \cdot)\) is not uniform.

Although the structure of the irreducible elements is more complicated, the sub-semigroups all have dimension 1, like the parent semigroup.

Theorem 16.4. Let \(k \in \mathbb{N}\). If \(\varphi\) is a homomorphism from \((S_k, \cdot)\) into \((\mathbb{R}, +)\), then \(\varphi = c\#\) for some \(c \in \mathbb{R}\). In particular, \(\dim(S_k, \cdot) = 1\).

Proof. Suppose that \(\varphi\) is a homomorphism from \((S_k, \cdot)\) into \((\mathbb{R}, +)\). We will show by induction on \(#(x)\) that \(\varphi(x) = c\#(x)\) where \(c = \varphi(\{k+1\})\). The result is trivially true if \(#(x) = 1\), since the only such \(x \in S_k\) is \(x = \{k+1\}\). Suppose now that \(\varphi(x) = c\#(x)\) for all \(x \in S_k\) with \(#(x) \leq n\). Let \(x \in S_k\) with \(#(x) = n+1\). If \(x\) is not an irreducible element, then \(x = uv\) where \(u, v \in S_k\), \(#(u) \leq n\), \(#(v) \leq n\), and \(#(u) + #(v) = #(x)\).

In this case,

\[
\varphi(x) = \varphi(u) + \varphi(v) = c\#(u) + c\#(v) = c\#(x)
\]

On the other hand, if \(x\) is irreducible, let \(j = \min\{i \in \mathbb{Z} : i+1 \not\in x\}\). Note that \(j < #(x)\) since \(\max(x) = #(x)+k\). Now let \(y \in S_k\) be obtained from \(x\) by replacing \(j\) with \(j+1\). Note that \(#(y) = #(x)\) and moreover, \(y^c\) can be obtained from \(x^c\) by replacing \(j+1\) with \(j\). We claim that \(xx = yx\): that is, \(x \cup (x^c \circ x) = y \cup (y^c \circ x)\).

To see this, note first that if \(i \neq j\) and \(i \neq j+1\), then \(i \in x\) if and only if \(i \in y\) and \(i \in x^c\) if and only if \(i \in y^c\). On the other hand, \(j \in x\) and \(j \in y^c \circ x\) since \(j = y^c[x(1)]\) (by definition, there are \(x(1)-1\) elements less than \(x(1)\) in \(y^c\); the next element in \(y^c\) is \(j\)). Similarly, \(j+1 \in y\) and \(j+1 \in x^c \circ x\) since \(j+1 = x^c[x(1)]\).

Since \(xx = yx\), it follows that \(\varphi(x) = \varphi(y)\). Continuing this process, we find that \(\varphi(x) = \varphi(y)\) for some \(y \in S_k\) that is not irreducible, but with \(#(y) = #(x)\). It then follows that \(\varphi(x) = \varphi(y) = c\#(y) = c\#(x)\) and the proof is complete.

Example 16.2. To illustrate the construction in the proof, let \(x = \{3, 4, 5, 8, 15\}\), so that \(x \in S_{10}\). Then \(j = 5\), \(y = \{3, 4, 6, 8, 15\}\) and

\[
xx = yx = \{3, 4, 5, 6, 7, 8, 9, 12, 15, 20\}
\]

For \(k \in \mathbb{N}\), the quotient space of \(S\) corresponding to positive subsemigroup \(T_k = S_k \cup \{0\}\) is

\[
S/T_k = \{z \in S : t \not\subseteq z\} \text{ for all } t \in M_k
\]

In particular, \(S/S_0 = \{z \in S : 1 \not\subseteq z\}\). In this case, the basic assumption for the decomposition of \(S\) over \(S_0\) and \(S/S_0\) is satisfied: For \(x \in S\) with \(1 \in x\) (so that \(x \not\subseteq S_0\)), let \(j_x = \max\{j \in \mathbb{N}_+ : \{1, 2, \ldots, j\} \subseteq x\}\). Then \(x\) can be factored uniquely as \(x = yz\) where

\[
y = \{1, 2, \ldots, j_x\} \in S_0
\]

\[
z = \{j - j_x : j \in x, j > j_x\} \in S/S_0
\]

If \(k \in \mathbb{N}_+\), then the basic assumption for the decomposition over \(T_k\) and \(S/T_k\) is not satisfied. For example, let \(x = \{1, 2, \ldots, k+2\} \in S_0\). The subsets of \(x\) that are in \(T_k\) are \(\{k+1\}\) and \(\{j, k+2\}\) for \(j \in \{1, 2, \ldots, k+1\}\). (Note that no subsets of \(x\) of cardinality greater than 2 are in \(T_k\).) So all of the sets \(y = \{j, k+2\}\) for \(j \in \{1, 2, \ldots, k+1\}\) are maximal elements of \(S_k\) with \(y \subseteq x\).

Since \((S_0, \cdot)\) is isomorphic to \((\mathbb{N}, +)\), the left walk function of order \(n \in \mathbb{N}\) for \((S_0, \subseteq)\) is

\[
\gamma_{0,n}(x) = \left(\frac{#(x)+n}{n}\right), \quad x \in S_0
\]

Problem 16.1. Find the left walk functions for \((S_k, \subseteq)\) when \(k \in \mathbb{N}_+\).

There are also sub-semigroups of \((S_k, \subseteq)\) based on the minimum.
Theorem 16.5. For $k \in \mathbb{N}_+$, let $U_k = \{x \in S : \min(x) = k\}$ and $V_k = \{x \in S : \min(x) \geq k\}$.

(a) $(U_k, \cdot)$ is a sub-semigroup of $(S, \cdot)$.

(b) $(V_k, \cdot)$ is a complete sub-semigroup of $(S, \cdot)$.

Proof. Trivially $U_k$ and $V_k$ are closed under $\cdot$ since $\min(xy) = \min\{\min(x), \min(y)\}$ for $x, y \in S$. So all that remains is to show that $(V_k, \cdot)$ is complete. Thus, suppose that $x, y \in V_k, z \in S$, and that $y = xz$. If $y = x$ then $z = \emptyset \in V_k$ (recall that $\min(\emptyset) = \infty$). Next, suppose that $y \neq x$ so that $x \subset y$ and $z \neq \emptyset$. Let $i = \min(z)$. If $i < k$ then, since $\min(x) \geq k$, $y$ contains the element $x^c(i) = i$. But this is a contradiction since $\min(y) \geq k$ so we must have $i \geq k$.

Clearly $(U_k, \cdot)$ is a sub-semigroup of $(V_k, \cdot)$ for $k \in \mathbb{N}_+$, and $(V_k, \cdot)$ is a sub-semigroup of $(V_j, \cdot)$ for $j, k \in \mathbb{N}_+$ with $j \leq k$. The collection $\{U_k : k \in \mathbb{N}_+\}$ is disjoint and partitions $S_+$ while the collection $\{V_k : k \in \mathbb{N}_+\}$ is decreasing with $V_1 = S_+$. More generally, $\{U_k : k \in \mathbb{N}_+, k \geq j\}$ partitions $V_j$ for $j \in \mathbb{N}_+$. The sub-semigroup $(U_k, \cdot)$ is not complete for $k \in \mathbb{N}_+$. For example, $\{k\}\{k+1\} = \{k, k+2\}$. Note that $\{k\}, \{k, k+2\} \in U_k$ but $\{k+1\} \notin U_k$. In fact, as the next theorem states, $\{k\}$ and $\{k, k+2\}$ are irreducible in $(U_k, \cdot)$.

Theorem 16.6. Let $k \in \mathbb{N}_+$.

(a) The set of irreducible elements of $(U_k, \cdot)$ is $\{x \in U_k : k+1 \notin x\}$.

(b) The set of irreducible elements of $(V_k, \cdot)$ is $\{x \in V_k : \#(x) = 1\} = \{\{k+j\} : j \in \mathbb{N}_+\}$.

Proof. Let $k \in \mathbb{N}_+$.

(a) First we show that if $y, z \in U_k$ then $k+1 \notin yz$. If $k+1 \notin y$ then $k+1 \notin yz$ since $y \subseteq yz$. Suppose that $z = y^{c}(k+1)$. Since $k$ is the minimum element of $y$, $k+1$ is the element of rank $k$ in $y^{c}$. But $k$ is also the minimum element of $z$ and hence $k+1 \notin yz$. Thus, if $x \in U_k$ is not irreducible then $k+1 \in x$.

Conversely, suppose that $x \in U_k$ and $k+1 \in x$. Note that if $y \in U_k$ then $\{k\}y = \{k\} \cup \{i+1 : i \in y\}$. Thus, define $y = \{i+1 : i \in x, i > k\}$. Then $y \in U_k$ and $x = \{k\}y$ so $x$ is not irreducible.

(b) Trivially, if $x \in V_k$ and $\#(x) = 1$ then $x$ is irreducible since it’s a singleton. Conversely, suppose that $x \in V_k$ and $\#(x) \geq 2$. Let $i = \min(x) \geq k$ and let $y = \{j-1 : j \in x, j \neq i\}$. Then just as in the proof of (a), $y \in V_k, \{i\} \in V_k$, and $x = \{i\}y$ so $x$ is not irreducible.

The sub-semigroup $(V_k, \cdot)$ is uniform for $k \in \mathbb{N}_+$. If $x \in V_k$ with $\#(x) = n \in \mathbb{N}_+$ then $x$ can be factored into irreducible elements. Since these are singletons, the number of factors must be $n$. On the other hand, the sub-semigroup $(U_k, \cdot)$ is not uniform for $k \in \mathbb{N}_+$. For example, $\{k\}$ and $\{k, k+2\}$ are irreducible in $(U_k, \cdot)$ but

$$\{k, k+2\}\{k\} = \{k\}^{3} = \{k, k+1, k+2\}$$

The next theorem shows that it suffices just to consider $(U_1, \cdot)$ and $(V_1, \cdot)$. Moreover, as noted earlier, $(V_1, \cdot)$ is the same as the strict positive semigroup $(S_+, \cdot)$ which we have already considered. So really, only $(U_1, \cdot)$ is new.

Theorem 16.7. Let $k \in \mathbb{N}_+$.

(a) $(U_k, \cdot)$ is isomorphic to $(U_1, \cdot)$.

(b) $(V_k, \cdot)$ is isomorphic to $(V_1, \cdot)$.

Proof. For $x \in S$ define $\phi_k(x) = \{i + (k-1) : i \in x\}$. It’s easy to see that $\phi_k$ is one-to-one. Moreover, $\phi_k$ maps $V_1$ onto $V_k$ and $U_1$ onto $U_k$. Finally, tracing through the definitions, we see that $\phi_k(x \cdot y) = \phi_k(x) \cdot \phi_k(y)$ for $x, y \in S$.

Let $\rightarrow$ denote the relation associated with $(U_1, \cdot)$, so that for $x, y \in U_1, x \rightarrow y$ if and only if $y = xz$ for some $z \in U_1$. Since $(U_1, \cdot)$ is not complete, $\rightarrow$ is not simply the subset relation. That is, for $x, y \in U_1, x \rightarrow y$ implies $x \subset y$, but not conversely. But $\rightarrow$ is easy to express with the help of a new function. Define $u : U_1 \rightarrow \mathbb{N}_+$ by

$$u(x) = \max\{k \in \mathbb{N}_+ : \{1, 2, \ldots, k\} \subseteq x\}, \quad x \in U_1$$

Note that $u(x)$ is well defined for $x \in U_1$ since $x$ is finite and $1 \in x$. 

Theorem 16.8. For \( x, y \in U_1 \), \( x \rightarrow y \) if and only if \( x \subseteq y \) and \( u(x) < u(y) \).

Proof. Suppose that \( x, y \in U_1 \) and \( x \rightarrow y \). Then by definition, \( y = xz \) for some \( z \in U_1 \). But then \( x \subseteq y \) (strict, since \( z \neq \emptyset \)), and since \( 1 \in z \), \( y \) contains the first element in \( x^c \), which by definition is \( u(x) + 1 \). Hence \( u(y) \geq u(x) + 1 \). Conversely, suppose that \( x, y \in U_1 \) and that \( x \subseteq y \) and \( u(x) < u(y) \). We know there exists \( z \in S \) such that \( xz = y \), so it just remains to show that \( 1 \in z \). But since \( u(y) \geq u(x) + 1 \), we must have \( u(x) + 1 \in y \). But this is the element of rank 1 in \( x^c \) so it follows that \( 1 \in z \).

The following theorem parallels Theorem 16.4.

Theorem 16.9. Suppose that \( \varphi \) is a homomorphism from \((U_1, \cdot)\) into \((\mathbb{R}, +)\). The \( \varphi = c\# \) for some \( c \in \mathbb{R} \). In particular, \( \dim(U_1, \cdot) = 1 \).

Proof. We prove that \( \varphi(x) = \varphi(\{1\})\#(x) \) for \( x \in U_1 \) by induction on \( \#(x) \). The result is trivial when \( \#(x) = 1 \) since \( x = \{1\} \). Suppose that the result holds for every \( x \in U_1 \) with \( \#(x) = n \), for some \( n \in \mathbb{N}_+ \). Suppose that \( x \in U_1 \) with \( \#(x) = n + 1 \). If \( 2 \in x \) then \( x = \{1\}y \) for some \( y \in U_1 \) with \( \#(y) = n \). But then by the homomorphism property and the induction hypothesis,

\[
\varphi(x) = \varphi(\{1\}) + \varphi(y) = \varphi(\{1\})(n + 1)
\]

Suppose \( 2 \notin x \) so that \( x \) is irreducible. Let \( y = \{1\} \cup \{i - 1 : i \in x, i > 1\} \). Then

\[
\{1\}x = y\{1\} = \{1\} \cup \{i + 1 : i \in x\}
\]

Note that \( \#(y) = \#(x) = n + 1 \) and from the homomorphism property \( \varphi(x) = \varphi(y) \). Repeating the procedure but with \( y \) replacing \( x \) we eventually have \( \varphi(x) = \varphi(z) \) for some \( z \in U_1 \) with \( 2 \in z \). But then just as before, \( z = \{1\}w \) with \( w \in U_1 \) and \( \#(w) = n \), so \( \varphi(x) = \varphi(\{1\})(n + 1) \).

The quotient space of \( S \) with respect to the positive sub-semigroup \( T := U_1 \cup \{\emptyset\} \) is

\[
S/T = \{z \in S : 1 \notin z\} = V_2
\]

The basic assumption for the decomposition of \( S \) over \( T \) and \( S/T \) is trivially satisfied, since if \( x \in S \) then either \( x \in T \) or \( x \in S/T \).

### 16.4 Exponential distributions

In this section, we explore the existence of exponential distributions for the semigroup \((S, \cdot)\) and the various sub-semigroups that we have defined. Here is the first main result:

Theorem 16.10. There are no memoryless distributions for \((S, \cdot)\), and hence no exponential distributions.

Proof. Suppose that \( X \) is a random variable with values in \( S \), so that \( X \) is a random, finite subset of \( \mathbb{N}_+ \). By our basic support assumption, note that \( \mathbb{P}(i \in X) = \mathbb{P}(\{i\} \subseteq X) > 0 \) for every \( i \in \mathbb{N}_+ \). Suppose that \( X \) has a memoryless distribution for \((S, \cdot)\). Then for every \( i \in \mathbb{N}_+ \) we have

\[
\mathbb{P}(\{i\} \subseteq X) = \mathbb{P}(i \in X) \mathbb{P}(i \in X)
\]

\[
\mathbb{P}(\{i + 1\} \subseteq X) = \mathbb{P}(i + 1 \in X) \mathbb{P}(i \in X)
\]

But \( \{i\} \{i\} = \{i + 1\} \{i\} = \{i, i + 1\} \) as noted in (16.1) and (16.2), so we must have

\[
\mathbb{P}(i + 1 \in X) = \mathbb{P}(i \in X)
\]

for every \( i \in \mathbb{N}_+ \). Next, note that if \( i_1, i_2, \ldots, i_n \in \mathbb{N}_+ \) with \( i_1 < i_2 < \cdots < i_n \) then by another application of the memoryless property,

\[
\mathbb{P}(i_1 \in X, i_2 \in X, \ldots, i_n \in X) = \mathbb{P}(\{i_1, i_2, \ldots, i_n\} \subseteq X)
\]

\[
= \mathbb{P}(\{i_n\} \{i_{n-1}\} \cdots \{i_1\} \subseteq X)
\]

\[
= \mathbb{P}(i_1 \in X) \mathbb{P}(i_2 \in X) \cdots \mathbb{P}(i_n \in X)
\]
It therefore follows that the events \( \{i \in X : i \in \mathbb{N}_+\} \) are independent with a common positive probability. By the Borel-Cantelli lemma, infinitely many of the events must occur with probability 1, so \( X \) is infinite—a contradiction.

Here’s another proof that there are no exponential distributions. Suppose that \( X \) has an exponential distribution for \((S, \cdot)\) with right probability function \( F \). Then \( X \) is memoryless so \( F(xy) = F(x)F(y) \) for \( x, y \in S \). But then \( \phi = \ln F \) is a homomorphism from \((S, \cdot)\) into \((\mathbb{R}, +)\) and so by Theorem \ref{thm:ln-F}, \( \ln F = -c\# \) for some \( c \in (0, \infty) \). Equivalently, \( F(x) = e^{-c\#(x)} \) for \( x \in S \). But then \( \sum_{x \in S} F(x) = \infty \) so \( F \) cannot be normalized into a probability density function.

Although there are no exponential distributions on \( S \), the sub-semigroups \((S_k, \cdot)\) has a one-parameter family of exponential distributions for each \( k \in \mathbb{N} \). The case \( k = 0 \) is trivial since \((S_0, \cdot)\) is isomorphic to \((\mathbb{N}, +)\). An exponential distribution for \((S_0, \cdot)\) has distribution function \( F \) and probability density function \( f \) of the form \( F(x) = \alpha \#(x) \) and \( f(x) = (1 - \alpha)\alpha \#(x) \) for \( x \in S_0 \), where \( \alpha \in (0, 1) \). Equivalently, \( \#(X) \) has the geometric distribution on \( \mathbb{N} \) with success parameter \( 1 - \alpha \).

**Theorem 16.11.** For \( k \in \mathbb{N}_+ \), a distribution is exponential for \((S_k, \cdot)\) if and only if the right probability function \( F \) and density function \( f \) have the following form, for some \( \alpha \in (0, 1) \):

\[
F(x) = \alpha \#(x), \quad x \in S_k \tag{16.7}
\]

\[
f(x) = (1 - \alpha)^{k+1} \alpha \#(x)^{-1}, \quad x \in S_k \tag{16.8}
\]

**Proof.** The function \( F(x) = \alpha \#(x) \) takes values in \((0, 1)\) and satisfies \( F(xy) = F(x)F(y) \) for all \( x, y \in S_k \), since \( \#(xy) = \#(x) + \#(y) \). Moreover

\[
\sum_{x \in S_k} F(x) = \sum_{n=1}^{\infty} \sum_{x \in S_{n,k}} F(x) = \sum_{n=1}^{\infty} \sum_{x \in S_{n,k}} \alpha^n = \sum_{n=1}^{\infty} \left( \frac{n+k-1}{n-1} \right) \alpha^n = \frac{\alpha}{(1 - \alpha)^{k+1}}
\]

It follows from Theorem 5.6 that \( F \) and \( f \) given in (16.7) and (16.8) are the right probability function and density function, respectively, of an exponential distribution with rate \((1 - \alpha)^{k+1}/\alpha\).

Conversely, suppose now that \( F \) is a right probability function on \( S_k \) with the memoryless property. Then \( \varphi = \ln F \) is a homomorphism from \((S_k, \cdot)\) into \((\mathbb{R}, +)\). By Theorem 16.4, \( \ln F = c\# \) where \( c = \ln[F(k + 1)] \). Hence \( F = \alpha \#(x) \) where \( \alpha = F(\{k + 1\}) \) \((0,1)\).

The density function of \( X \) depends on \( x \in S_k \) only through \( \#(x) \), so it is natural to study the distribution of \( \#(X) \). Of course, by definition \( \max(X) = \#(X) + k \) on \( S_k \), so the distribution of \( \#(X) \) determines the distribution of \( \max(X) \). Once again, the case \( k = 0 \) is trivial. If \( X \) has the exponential distribution on \((S_0, \cdot)\) then \( \#(X) \) has the geometric distribution on \( \mathbb{N} \) with success parameter \( \alpha \).

**Corollary 16.3.** Suppose that \( k \in \mathbb{N}_+ \) and that \( X \) has the exponential distribution on \((S_k, \cdot)\) with parameter \( \alpha \in (0, 1) \). Then

\[
\mathbb{P}[\#(X) = n] = \binom{n+k-1}{k}(1 - \alpha)^{k+1} \alpha^{n-1}, \quad n \in \mathbb{N}_+
\]

\[
\mathbb{E}[\#(X)] = \frac{1 + k\alpha}{1 - \alpha}
\]

\[
\text{var}[\#(X)] = \frac{(k+1)\alpha}{(1 - \alpha)^2}
\]

**Proof.** Note that \( \#(X) = n \) if and only if \( X \in S_{n,k} \). Hence the formula for the probability density function follows from Theorem 16.11 and the fact that \( \#(S_{n,k}) = \binom{n+k-1}{k} \). Note next that \( \#(X) - 1 \) has the negative binomial distribution on \( \mathbb{N} \) with stopping parameter \( k+1 \) and success parameter \( 1 - \alpha \). The formulas for the mean and variance then follow from standard results.

**Corollary 16.4.** Suppose that \( k \in \mathbb{N} \) and that \( X \) has the exponential distribution on \((S_k, \cdot)\) with parameter \( \alpha \in (0, 1) \). Given \( \#(X) = n \in \mathbb{N} \), \( X \) is uniformly distributed on \( S_{n,k} \).
Proof. The case $n = 0$ is trivial since $\#(X) = 0$ if and only if $X = \emptyset$, and in this case, $k = 0$ also. Suppose $n \in \mathbb{N}_+$. Using the probability density functions of $X$ and $\#(X)$ given above we have

$$
P[X = x \mid \#(X) = n] = \frac{\mathbb{P}(X = x)}{\mathbb{P}[\#(X) = n]} = \frac{1}{(n + k - 1)}; \quad x \in S_{n,k}
$$

It is easy to see from the Corollary 16.3 that for each $k \in \mathbb{N}_+$, $\mathbb{E}[\#(X)]$ is a strictly increasing function of $\alpha$ and maps $(0, 1)$ onto $(1, \infty)$. Thus, the exponential distribution on $S_k$ can be re-parameterized by expected cardinality. Moreover, the exponential distribution maximizes entropy with respect to this parameter:

**Corollary 16.5.** Suppose that $k \in \mathbb{N}$ and that $X$ has the exponential distribution on $(S_k, \cdot)$ with parameter $\alpha \in (0, 1)$. Then $X$ maximizes entropy over all random variables $Y$ on $S_k$ with

$$
\mathbb{E}[\#(Y)] = \mathbb{E}[\#(X)] = \frac{1 + k\alpha}{1 - \alpha}
$$

Proof. We use the usual inequality for entropy in Theorem 4.6: if $f$ and $g$ are probability density functions on $X$ and $Y$, respectively, then

$$
- \sum_{x \in S_k} g(x) \ln g(x) \leq - \sum_{x \in S_k} g(x) \ln f(x)
$$

(16.9)

If $X$ has the exponential distribution with parameter $\alpha$, and $\mathbb{E}[\#(Y)] = \mathbb{E}[\#(X)]$ then substituting into the right side of equation (16.9) we see that the entropy of $Y$ is bounded above by

$$(k + 1) \ln(1 - \alpha) + \ln \alpha - \ln(\alpha) \frac{1 + k\alpha}{1 - \alpha}$$

Of course, the entropy of $X$ achieves this upper bound.

**Problem 16.2.** Suppose that $k \in \mathbb{N}_+$ and that $Y = (Y_1, Y_2, \ldots)$ is the random walk on $(S_k, \cdot)$ associated with the exponential distribution with parameter $\alpha \in (0, 1)$. Find a closed form expression for the density function $f_n$ of $Y_n$ for $n \in \{2, 3, \ldots\}$. This is equivalent to the Problem 16.1 of finding the left walk function of order $n - 1$ for the graph $(S_k, \subset)$.

**Problem 16.3.** For $k \in \mathbb{N}_+$, determine whether the exponential distribution on $(S_k, \cdot)$ with parameter $\alpha \in (0, 1)$ is compound Poisson (and therefore infinitely divisible relative to the semigroup operation).

**Problem 16.4.** Suppose that $X$ has an exponential distribution on $(S_k, \cdot)$ with parameter $\alpha \in (0, 1)$. Compute the hitting probability function on $S$. That is, compute $\mathbb{P}(X \cap x \neq \emptyset)$ for $x \in S$.

For $k \in \mathbb{N}_+$, the exponential distributions for $(S_k, \cdot)$ also have constant rate for the various relations associated with the semigroup: the strict partial order $\subset$, the partial order $\subseteq$, the covering relation $\uparrow$, and the reflexive completion $\uparrow$ of $\uparrow$.

**Theorem 16.12.** Suppose that $k \in \mathbb{N}_+$ and that $X$ has the exponential distribution on $(S_k, \cdot)$ with parameter $\alpha \in (0, 1)$.

(a) For the graph $(S_k, \subset)$, $X$ has constant rate

$$
\frac{(1 - \alpha)^{k+1}}{\alpha}
$$

(b) For the graph $(S_k, \subseteq)$, $X$ has constant rate

$$
\frac{(1 - \alpha)^{k+1}}{\alpha + (1 - \alpha)^{k+1}}
$$

(c) For the graph $(S_k, \uparrow)$, $X$ has constant rate

$$
\frac{1}{\alpha(1 + \alpha)^k}
$$
(d) For the graph \((S_k, \uparrow)\), \(X\) has constant rate

\[
\frac{1}{\alpha(1+\alpha)^k + 1}
\]

Proof. The proofs are straightforward.

(a) This follows immediately from the Theorem 16.11 since the relation associated with the strict positive semigroup \((S_k, \cdot)\) is the strict partial order \(\subset\).

(b) This follows from (a) and standard results for reflexive completion.

(c) As before, let \(M_k\) denote the set of irreducible elements of \((S_k, \cdot)\) and let \(M_{nk} = \{y \in M_k : \#(y) = n\}\) for \(n \in \{1, 2, \ldots, k+1\}\). As noted in the proof of Theorem 16.3, \(M_{nk} : n \in \{1, 2, \ldots, k+1\}\) partitions \(M_k\) and \(\#(M_{nk}) = \binom{k}{n-1}\). So the right probability function \(G\) of \(X\) for \((S_k, \uparrow)\) is given by

\[
G(x) = \sum_{y \in M_k} f(x \cdot y) = \sum_{n=1}^{k+1} \sum_{y \in M_{nk}} f(x \cdot y) = \sum_{n=1}^{k+1} (1-\alpha)^{k+1} \alpha^{\#(x)+n-1} \binom{k}{n-1}
\]

\[
= f(x) \sum_{n=1}^{k+1} \alpha^n \binom{k}{n-1} = f(x)\alpha(1+\alpha)^k
\]

(d) This follows from (c) and standard results for reflexive completion.

We now turn our attention to the sub-semigroup \((U_k, \cdot)\) where \(U_k = \{x \in S : \text{min}(x) = k\}\) for \(k \in \mathbb{N}_+\). Recall that \((U_k, \cdot)\) is isomorphic to \((U_1, \cdot)\) for \(k \in \mathbb{N}_+\), so we really only need to consider \((U_1, \cdot)\).

**Theorem 16.13.** There are no exponential distributions for the semigroup \((U_1, \cdot)\) (and hence no exponential distributions for the semigroup \((U_k, \cdot)\) for \(k \in \mathbb{N}_+\)).

**Proof.** Suppose that \(X\) is a random variable with values in \(U_1\) and right probability function \(F\) with respect to \((U_1, \cdot)\). If \(X\) has an exponential distribution, then the distribution is memoryless and so \(\ln F\) is a homomorphism from \((U_1, \cdot)\) into \((\mathbb{R}, +)\). By Theorem 16.9, \(F(x) = |F(\{1\})|^{\#(x)}\) for \(x \in U_1\). Let \(U_{1n} = \{x \in U_1 : \#(x) = n\}\) for \(n \in \mathbb{N}_+\). Then

\[
\sum_{x \in U_1} F(x) = \sum_{n=1}^{\infty} \sum_{x \in U_{1n}} F(x) = \sum_{n=1}^{\infty} |F(\{1\})|^n \#(U_{1n}) = \infty
\]

since \(\#(U_{1n}) = \infty\) for \(n \in \{2, 3, \ldots\}\). Hence \(F\) cannot be normalized into a probability density function.

### 16.5 Almost exponential distributions

There are no exponential distributions on \((S, \cdot)\). However, we can define distributions that are “close” to exponential by forming mixtures of distributions that are exponential for \((S_k, \cdot)\) over \(k \in \mathbb{N}\). Thus, suppose that \(X\) takes values in \(S\) with probability density function \(f\) given by

\[
f(x) = \begin{cases} 
p_0(1-\alpha_0)\alpha_0^{\#(x)}, & x \in S_0 
p_k(1-\alpha_k)^{k+1}\alpha_k^{\#(x)-1}, & x \in S_k, \ k \in \mathbb{N}_+ \end{cases}
\]

(16.10)

where \(\alpha_k, p_k \in (0, 1)\) for each \(k \in \mathbb{N}\) and \(\sum_{k=0}^{\infty} p_k = 1\). Thus, the conditional distribution of \(X\) given \(X \in S_k\) is exponential on \((S_k, \cdot)\), with parameter \(\alpha_k\), for each \(k \in \mathbb{N}\). The distribution of \(X\) is as close to exponential as possible, in the sense that \(X\) is essentially exponential on each of the sub-semigroups \(S_k\), and these semigroups partition \(S\). We can construct random variable \(X\) as follows: For \(k \in \mathbb{N}_+\), suppose that \(Y_k\) has the exponential distribution on \(S_k\) with parameter \(\alpha_k \in (0, 1)\). Next, suppose that \(K\) takes values in \(\mathbb{N}\) with probability density function given by \(P(K = k) = p_k\) for \(k \in \mathbb{N}\). Then \(X = Y_K\) has the probability density function given in (16.10).
There is not much that we can say about the general distribution defined by (16.10). In the remainder of this section we will study a special case with particularly nice properties. For our first construction, we define a random variable \( X \) on \( S \) by first selecting a geometrically distributed population size \( N \), and then selecting a sample from \( \{1, 2, \ldots, N\} \) in an IID fashion. Of course, geometric distributions are the exponential distributions for the positive semigroup \((\mathbb{N}, +)\). More precisely, let \( N \) be a random variable taking values in \( \mathbb{N} \), and having the geometric distribution with success parameter \( 1 - r \) where \( r \in (0, 1) \):

\[
P(N = n) = (1 - r)r^n, \quad n \in \mathbb{N}
\]

Next, given \( N = n \), \( X \) is distributed on the subsets of \( \{1, 2, \ldots, n\} \) so that \( i \in X \), independently, with probability \( p \) for each \( i \in \{1, 2, \ldots, n\} \). Of course, if \( N = 0 \), then \( X = \emptyset \).

**Theorem 16.14.** Random variable \( X \) has probability density function \( f \) and right probability function \( F \) given by

\[
f(x) = \frac{1 - r}{1 - r(1 - p)}(rp)^{\#(x)}[r(1 - p)]^{\max(x) - \#(x)}, \quad x \in S \tag{16.11}
\]

\[
F(x) = p^{\#(x)}r^{\max(x)}, \quad x \in S \tag{16.12}
\]

**Proof.** For \( x \in S \),

\[
P(X = x) = \sum_{n=0}^{\infty} P(N = n)P(X = x \mid N = n)
\]

If \( n < \max(x) \) then \( x \) is not a subset of \( \{1, 2, \ldots, n\} \), so \( P(X = x \mid N = n) = 0 \). If \( n \geq \max(x) \) then \( x \) is a subset of \( \{1, 2, \ldots, n\} \) and by assumption, \( P(X = x \mid N = n) = p^{\#(x)}(1 - p)^{n - \#(x)} \). Substituting gives

\[
P(X = x) = \sum_{n=\max(x)}^{\infty} (1 - r)r^n p^{\#(x)}(1 - p)^{n - \#(x)}
\]

which simplifies to (16.11). By a similar argument,

\[
P(X \geq x) = \sum_{n=\max(x)}^{\infty} (1 - r)r^n p^{\#(x)}
\]

which simplifies to (16.12).

Not surprisingly, the distribution of \( X \) depends on \( x \in S \) only through \( \#(x) \) and \( \max(x) - \#(x) \). As before, let \( U = \#(X) \) and now let \( V = \max(X) - \#(X) \). The following corollaries will explore the relationships between the distributions of \( U, V, \) and \( X \), and provide another way of constructing the distribution of \( X \).

**Corollary 16.6.** The joint probability density function of \((U, V)\) is given by

\[
P(U = 0, V = 0) = \frac{1 - r}{1 - r(1 - p)}
\]

\[
P(U = n, V = k) = \frac{1 - r}{1 - r(1 - p)} \binom{n + k - 1}{n - 1} (rp)^n[r(1 - p)]^k, \quad n \in \mathbb{N}_+, k \in \mathbb{N}
\]

**Proof.** Note that \( \{U = n, V = k\} = \{X \in S_{n,k}\} \). Recall that \( S_{0,0} = \{\emptyset\} \) and \( \#(S_{n,k}) = \binom{n+k-1}{n-1} \) if \( n \in \mathbb{N}_+, k \in \mathbb{N} \). So the result follows from (16.11).

**Corollary 16.7.** For \( n = k = 0 \) or \( n \in \mathbb{N}_+, k \in \mathbb{N} \), the conditional distribution of \( X \) given \( U = n, V = k \) is uniform on \( S_{n,k} \).

**Proof.** First the trivial case: \( P(X = \emptyset \mid U = 0, V = 0) = 1 \). If \( n \in \mathbb{N}_+, k \in \mathbb{N} \) then from the density functions of \( X \) and \((U, V)\),

\[
P(X = x \mid U = n, V = k) = \frac{P(X = x)}{P(U = n, V = k)} = \frac{1}{\binom{n+k-1}{n-1}}, \quad x \in S_{n,k}
\]
Corollary 16.8. U has a geometric distribution on \( \mathbb{N} \) with success parameter \( (1 - r)/(1 - r(1 - p)) \):

\[
\mathbb{P}(U = n) = \frac{1 - r}{1 - r(1 - p)} \left[ \frac{rp}{1 - r(1 - p)} \right]^n, \quad n \in \mathbb{N}
\]

\[
\mathbb{E}(U) = \frac{rp}{1 - r}, \quad \text{var}(U) = \frac{rp[1 - r(1 - p)]}{(1 - r)^2}
\]

Proof. First,

\[
\mathbb{P}(U = 0) = \mathbb{P}(U = 0, V = 0) = \frac{1 - r}{1 - r(1 - p)}
\]

Next using the general binomial theorem,

\[
\mathbb{P}(U = n) = \sum_{k=0}^{\infty} \mathbb{P}(U = n, V = k) = \frac{1 - r}{1 - r(1 - p)} \left( \frac{r p}{1 - r(1 - p)} \right)^n \sum_{k=0}^{\infty} {n + k - 1 \choose k} [r(1 - p)]^k
\]

\[
= \frac{1 - r}{1 - r(1 - p)} \left[ \frac{rp}{1 - r(1 - p)} \right]^n, \quad n \in \mathbb{N}_+
\]

Corollary 16.9. For \( n \in \mathbb{N}_+ \), the conditional distribution of \( V \) given \( U = n \) is negative binomial on \( \mathbb{N} \) with stopping parameter \( n \) and success parameter \( 1 - r(1 - p) \):

\[
\mathbb{P}(V = k \mid U = n) = {n + k - 1 \choose n - 1} [r(1 - p)]^k [1 - r(1 - p)]^n, \quad k \in \mathbb{N}
\]

\[
\mathbb{E}(V \mid U = n) = n \frac{r(1 - p)}{1 - r(1 - p)}, \quad \text{var}(V \mid U = n) = n \frac{r(1 - p)}{[1 - r(1 - p)]^2}
\]

Proof. This follows easily from the joint density of \((U, V)\) and the density of \( U \): For \( n \in \mathbb{N}_+ \),

\[
\mathbb{P}(V = k \mid U = n) = \frac{\mathbb{P}(U = n, V = k)}{\mathbb{P}(U = n)}, \quad k \in \mathbb{N}
\]

Of course also, \( P(V = 0 \mid U = 0) = 1. \)

Corollary 16.10. V has a modified geometric distribution on \( \mathbb{N} \):

\[
\mathbb{P}(V = 0) = \frac{1 - r}{[1 - r(1 - p)](1 - rp)}
\]

\[
\mathbb{P}(V = k) = \frac{(1 - r)rp}{[1 - r(1 - p)](1 - rp)} \left[ \frac{r(1 - p)}{1 - rp} \right]^k, \quad k \in \mathbb{N}_+
\]

\[
\mathbb{E}(V) = \frac{r^2p(1 - p)}{(1 - r)[1 - r(1 - p)]}, \quad \text{var}(V) = \frac{r^2(1 - p)[1 - rp(1 - p)]}{(1 - r)^2[1 - r(1 - p)]^2}
\]

Proof. This follows from the joint density function of \((U, V)\):

\[
\mathbb{P}(V = 0) = \sum_{n=0}^{\infty} \mathbb{P}(U = n, V = 0)
\]

and

\[
\mathbb{P}(V = k) = \sum_{n=1}^{\infty} \mathbb{P}(U = n, V = k), \quad k \in \mathbb{N}_+
\]

The moment results can also be obtained by conditioning:

\[
\mathbb{E}(V) = \mathbb{E}[\mathbb{E}(V \mid U)], \quad \text{var}(V) = \text{var}[\mathbb{E}(V \mid U)] + \mathbb{E}[\text{var}(V \mid U)]
\]
Corollary 16.11. The covariance and correlation of \((U, V)\) are
\[
\text{cov}(U, V) = \frac{r^2 p(1 - p)}{(1 - r)^2}
\]
\[
\text{cor}(U, V) = \sqrt{\frac{r p(1 - p)[1 - r(1 - p)]}{1 - r p(1 - p)}}
\]

Proof. Using the conditional mean of \(V\) given \(U\) we have
\[
E(UV) = E[E(UV \mid U)] = E(U E(V \mid U)] = \frac{r(1 - p)}{1 - r(1 - p)} E(U^2)
\]
But of course \(E(U^2) = \text{var}(U) + [E(U)]^2\) and \(\text{cov}(U, V) = E(UV) - E(U) E(V)\) so the covariance result follows from substituting the moments of \(U\) and \(V\) and simplifying.

Not surprisingly, \(U\) and \(V\) are positively correlated, but only weakly. The maximum correlation is about 0.37.

Of course, the corollaries determine the distribution of \(X\). In fact these results give an alternate way of constructing the distribution of \(X\) in the first place: We first give \(U\) a geometric distribution with success parameter \(b \in (0, 1)\); given \(U = 0\) set \(V = 0\) but otherwise given \(U = n \in \mathbb{N}\) we give \(V\) a negative binomial distribution on \(\mathbb{N}\) with stopping parameter \(n\) and success parameter \(b \in (0, 1)\); and finally, given \(U = n, V = k\) (so either \(n = k = 0\), or \(n \in \mathbb{N}\) and \(k \in \mathbb{N}\)), we give \(X\) the uniform distribution on \(S_{n,k}\).

Our original construction, although simple, is perhaps unsatisfactory because the population variable \(N\) is hidden (not directly observable from \(X\)). The alternate construction has no hidden variables, and moreover, the geometric distribution of \(U\) and the conditional uniform distribution for \(X\) given \(U = n, V = k\) are natural. On the other hand, the conditional negative binomial distribution of \(V\) given \(U = n\) is somewhat obscure. The two constructions are equivalent, since there is a one-to-one correspondence between the pairs of parameters:

\[
a = \frac{r p}{1 - r(1 - p)} \quad \quad b = r(1 - p)
\]
\[
r = a(1 - b) + b \quad \quad p = \frac{a(1 - b)}{a(1 - b) + b}
\]

Our next goal is to study the distribution of the random subset \(X\) on the sub-semigroups \(S_k\). First note that rate function of \(X\) on \((S, \cdot)\) is given by
\[
P(X = x) = \frac{1 - r}{1 - r(1 - p)} (1 - p)^{\text{max}(x) - \#(x)} \quad x \in S
\]
Thus, for \(k \in \mathbb{N}\), \(X\) has constant rate \(\frac{1 - r}{1 - r(1 - p)} (1 - p)^k\) on \(S_k\). The rate decreases to 0 exponentially fast as \(\text{max}(x) - \#(x)\) increases to \(\infty\). In particular, for \(x \in S_0\),
\[
P(X = x) = \frac{1 - r}{1 - r + r p} (r p)^{\#(x)}
\]
\[
P(X \geq x) = (r p)^{\#(x)}
\]
Hence, \(X\) has the memoryless property on \(S_0\), in addition to the constant rate property. That is, for \(x, y \in S_0\),
\[
P(X \geq xy) = (r p)^{\#(xy)} = (r p)^{\#(x)+\#(y)}
\]
\[
= (r p)^{\#(x)}(r p)^{\#(y)} = P(X \geq x)P(X \geq y)
\]
Next we find the conditional distribution of \(X\) given \(X \in S_k\) for \(k \in \mathbb{N}\).

Corollary 16.12. The conditional probability density functions of \(X\) give \(X \in S_k\) are as follows:
\[
P(X = x \mid X \in S_0) = (1 - r p)(r p)^{\#(x)} \quad x \in S_0
\]
\[
P(X = x \mid X \in S_k) = (1 - r p)^{k+1}(r p)^{\#(x) - 1} \quad x \in S_k, k \in \mathbb{N}_+
\]
Proof. This follows from the probability density functions of $X$ and $V$. For $k \in \mathbb{N}$,

$$P(X = x \mid X \in S_k) = \frac{P(X = x)}{P(X \in S_k)} = \frac{P(X = x)}{P(V = k)}, \quad x \in S_k$$

Thus, $X$ has an almost exponential distribution in the sense of the original definition in (16.11), with parameters $\alpha_k = rp$ for $k \in \mathbb{N}$, and with the mixing probabilities given by the probability density function of $V$ in Corollary 16.10.

Recall that no exponential distribution on $S$ exists because otherwise the events $\{\{i \in X\} : i \in \mathbb{N}_+\}$ would be independent with a common probability. The next corollary explores these events for the random variable constructed above.

**Corollary 16.13.** Suppose that $X$ has the almost exponential distribution with parameters $r, p \in (0, 1)$.

(a) $P(i \in X) = pr^i$ for $i \in \mathbb{N}_+$.

(b) If $i_1, i_2, \ldots, i_n \in \mathbb{N}_+$ with $i_1 < i_2 < \cdots < i_n$ then

$$P(i_n \in X \mid i_1 \in X, \ldots, i_{n-1} \in X) = P(i_n \in X \mid i_{n-1} \in X) = P(i_n - i_{n-1} \in X) = pr^{i_n - i_{n-1}}$$

(c) For $j \in \mathbb{N}_+$, the events $\{1 \in X\}, \{2 \in X\}, \ldots, \{j - 1 \in X\}$ are conditionally independent given $\{j \in X\}$ with $P(i \in X \mid j \in X) = p$ for $i \in \{1, 2, \ldots, j - 1\}$.

Proof. These results follow from the right probability function of $X$ for $(S, \subseteq)$ given in (16.12).

(a) $P(i \in X) = P(\{i\} \subseteq X) = pr^i$ for $i \in \mathbb{N}_+$.

(b) For $i_1, i_2, \ldots, i_n \in \mathbb{N}_+$ with $i_1 < i_2 < \cdots < i_n$,

$$P(i_n \in X \mid i_1 \in X, \ldots, i_{n-1} \in X) = \frac{P(\{i_1, \ldots, i_n\} \subseteq X)}{P(\{i_1, \ldots, i_{n-1}\} \subseteq X)} = \frac{p^n r^{i_n}}{p^{n-1} r^{i_{n-1}}} = pr^{i_n - i_{n-1}}$$

Similarly,

$$P(i_n \in X \mid i_{n-1} \in X) = \frac{P(\{i_{n-1}, i_n\} \subseteq X)}{P(\{i_{n-1}\} \subseteq X)} = \frac{p^2 r^{i_n}}{pr^{i_{n-1}}} = pr^{i_n - i_{n-1}}$$

and the common values is $P(i_n - i_{n-1} \in X)$ by part (a).

(c) First note that if $i \in \{1, 2, \ldots, j - 1\}$ then

$$P(i \in X \mid j \in X) = \frac{P(\{i, j\} \subseteq X)}{P(\{j\} \subseteq X)} = \frac{p^2 r^j}{pr^j} = p$$

Next, if $A \subseteq \{1, 2, \ldots, j - 1\}$ with $\#(A) = n$ then

$$P(A \subseteq X \mid j \in X) = \frac{P(A \cup \{j\} \subseteq X)}{P(\{j\} \subseteq X)} = \frac{p^{n+1} r^j}{pr^j} = p^n$$

Property (a) is clearly a result of the original construction of $X$. Property (b) is reminiscent of a Markov property. This property implies that the events $\{i \in X\} : i \in \mathbb{N}_+$ are positively correlated, but asymptotically uncorrelated. In fact the correlation decays exponentially since

$$P(i + j \in X \mid i \in X) = P(j \in X) = pr^j \to 0 \text{ as } j \to \infty$$

**Problem 16.5.** Characterize all random subsets of $\mathbb{N}_+$ that satisfy the “partial Markov property” above.

The proof of the following corollary is essentially the same as the proof of Corollary 16.5.
Corollary 16.14. Suppose that $X$ has the almost exponential distribution with parameters $p, r \in (0,1)$. Then $X$ maximizes entropy among all random variables $Y$ on $S$ with

$$
E[\#(Y)] = E[\#(X)] = \frac{rp}{1-r}
$$

$$
E[\max(Y)] = E[\max(X)] = \frac{rp}{(1-r)[1-r(1-p)]}
$$

Of fundamental importance in the general theory of random sets (see Matheron [25]) is the hitting probability function $G$:

$$
G(x) = P(X \cap x \neq \emptyset), \quad x \subseteq \mathbb{N}_+
$$

This function completely determines the distribution of a random set. Note that $G$ is defined for all subsets of the positive integers, not just finite subsets.

Theorem 16.15. Suppose that $X$ has the almost exponential distribution with parameters $p, r \in (0,1)$. Then

$$
G(x) = \sum_{i=1}^{\#(x)} p(1-p)^{i-1}r^{x(i)}, \quad x \subseteq \mathbb{N}_+
$$

where as usual, $x(i)$ is the $i$th smallest element of $x$.

Proof. Suppose first that $x$ is finite (so that $x \in S$). From the standard inclusion-exclusion formula (or from [25]),

$$
G(x) = \sum_{k=1}^{\#(x)} (-1)^{k-1} \sum_{y \subseteq x, \#(y)=k} F(y)
$$

Hence, substituting the formula for $F(y)$ we have

$$
G(x) = \sum_{k=1}^{\#(x)} (-1)^{k-1} \sum_{y \subseteq x, \#(y)=k} p^{\#(y)}r^{\max(y)}
$$

$$
= \sum_{k=1}^{\#(x)} (-1)^{k-1}p^k \sum_{i=k}^{\#(x)} \{r^{x(i)} : y \subseteq x, \#(y) = k, \max(y) = x(i)\}
$$

$$
= \sum_{k=1}^{\#(x)} (-1)^{k-1}p^k \sum_{i=k}^{\#(x)} \binom{i-1}{k-1} r^{x(i)} = \sum_{i=1}^{\#(x)} r^{x(i)} \sum_{k=1}^{i} \binom{i-1}{k-1} (-1)^{k-1}p^k
$$

$$
= \sum_{i=1}^{\#(x)} p(1-p)^{i-1}r^{x(i)}
$$

For infinite $x$, the formula holds by the continuity theorem.

Example 16.3. An easy computation gives

$$
G(\{n, n+1, \ldots\}) = \frac{pr^n}{1-r + rp}, \quad n \in \mathbb{N}_+
$$

This is also $P(\max(X) \geq n)$, and agrees with our earlier results. In particular, letting $n = 1$, we get $G(\mathbb{N}_+) = P(X \neq \emptyset)$:

$$
G(\mathbb{N}_+) = \frac{rp}{1-r + rp}
$$

This agrees with $1 - P(X = \emptyset)$ using the probability density function of $X$.

Exercise 16.7. Find the probability that $X$ contains an even integer.

Exercise 16.8. Find the probability that $X$ contains an odd integer.

Example 16.4. The following list gives 10 simulations of $X$ with $r = 0.95$ and $p = 0.65$, corresponding to $E[\#(X)] = 12.35$ and $E[\max(X)] = 18.5$. 
1. \( x = \{1, 2, 3\} \)
2. \( x = \{1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 16, 17, 18, 19, 20, 21, 22, 25, 26, 27, 32\} \)
3. \( x = \{2, 3, 5\} \)
4. \( x = \{1, 4, 5, 9, 10, 11, 12\} \)
5. \( x = \{2, 5, 6, 7, 8, 12, 13, 15, 17, 18, 20, 23, 24, 25, 26, 27, 29, 30, 31, 32, 35, 37, 38, 40, 41\} \)
6. \( x = \{1, 2, 6, 8, 10, 12, 15, 16, 17, 18, 20, 23, 25, 28, 29, 30, 31, 32, 34\} \)
7. \( x = \{1, 2, 3\} \)
8. \( x = \{1, 3, 4, 7, 10, 11, 12, 13, 14, 21, 30, 32, 33, 34, 35, 37, 38, 40, 41, 42, 45, 46, 47, 49, 50, 51, 52, 53\} \)
9. \( x = \{1, 4, 5, 6, 9\} \)
10. \( x = \{5, 7\} \)

**Problem 16.6.** Determine how the random sets in this section relate to the random closed sets studied by Matheron [25] and others.

**Problem 16.7.** Suppose that \( X = (X_1, X_2, \ldots) \) is a sequence of independent variables, each having the almost exponential distribution with parameters \( r, p \in (0, 1) \). Let \( Y_n = X_1 X_2 \cdots X_n \) for \( n \in \mathbb{N}_+ \) so that \( Y = (Y_1, Y_2, \ldots) \) is the random walk on \((S, \cdot)\) corresponding to the distribution. Find the density function of \( Y_n \).

Our last discussion concerns the distribution that is almost exponential for the semigroup \((S, \cdot)\) but now relative to the covering graph \((S, \uparrow)\). As usual, let \( v(x) = \max(x) - \#(x) \) for \( x \in S \).

**Theorem 16.16.** Suppose again that \( X \) has the almost exponential distribution on \((S, \cdot)\) with parameters \( r, p \in (0, 1) \). The right probability function \( H \) of \( X \) for \((S, \uparrow)\) is given by

\[
H(x) = \frac{1 - r}{1 - r(1 - p)} (rp)^{\#(x) + 1} [r(1 - p)]^{v(x) - 1} 
\left[ v(x) + \frac{r(1 - p)}{1 - r(1 - p)} \right], \quad x \in S
\]

**Proof.** Since \( \uparrow \) is the covering relation and since \( i > \max(x) \) implies \( i \notin x \), we have

\[
H(x) = \sum_{i \notin x} f(x \cup \{i\}) = \sum_{i \notin x, i < \max(x)} f(x \cup \{i\}) + \sum_{i = \max(x) + 1} f(x \cup \{i\})
\]

If \( i \notin x \) then \( \#(x \cup \{i\}) = \#(x) + 1 \). If in addition \( i < \max(x) \) then \( \max(x \cup \{i\}) = \max(x) \) so \( v(x \cup \{i\}) = v(x) - 1 \). Moreover, the number of terms in the first sum on the right is \( v(x) \) and so the first sum is

\[
v(x) \cdot \frac{1 - r}{1 - r(1 - p)} (rp)^{\#(x) + 1} [r(1 - p)]^{v(x) - 1}
\]

On the other hand, if \( i > \max(x) \) then \( \max(x \cup \{i\}) = i \) and so \( v(x \cup \{i\}) = i - \#(x) - 1 \). The second sum on the right is

\[
\sum_{i = \max(x) + 1}^{\infty} \frac{1 - r}{1 - r(1 - p)} (rp)^{\#(x) + 1} [r(1 - p)]^{i - \#(x) - 1} = \frac{1 - r}{1 - r(1 - p)^2} (rp)^{\#(x) + 1} [r(1 - p)]^{v(x)}
\]

simplifying gives the result. \( \square \)

We can rewrite the result as

\[
H(x) = f(x) rp \left[ \frac{v(x)}{r(1 - p)} + \frac{1}{1 - r(1 - p)} \right], \quad x \in S
\]

so in particular, on \( S_k \) the right rate function for \((S, \uparrow)\) is the constant

\[
r(1 - p)[1 - r(1 - p)]
\]

\[
\frac{r(1 - p)k \{k + r(1 - p)\}}{rp(k - r(1 - p))}
\]
16.6 Constant rate distributions

Recall that \((S, \cdot)\) does not have any exponential distributions. But does does the partial order graph \((S, \subseteq)\) have constant rate distributions? Similarly, does the covering graph \((S, \uparrow)\) have constant rate distributions?. First we note some simple results for the right probability functions. Let \(X\) be a random variable with values in \(S\) with probability density function \(f\).

**Theorem 16.17.** Let \(F\) denote the right probability function of \(X\) for \((S, \subseteq)\) so that

\[
F(x) = \sum_{y \in x S} f(y) = \sum_{x \subseteq y} f(y), \quad x \in S
\]

If \(\mathbb{E}[2^\#(X)] < \infty\) then

\[
f(x) = \sum_{x \subseteq y} (-1)^{\#(y) - \#(x)} F(y), \quad x \in S
\]

**Proof.** This follows immediately from the Möbius inversion result in Corollary 4.4 and the Möbius function and walk function given earlier.

**Theorem 16.18.** The right probability functions for the other graphs are as follows:

(a) For the graph \((S, \subset)\),

\[
F(x) = \sum_{y \in x S_+} f(y) = \sum_{x \subset y} f(y), \quad x \in S
\]

(b) For the graph \((S, \uparrow)\),

\[
F(x) = \sum_{i=1}^{\infty} f(x\{i\}) = \sum_{j \notin x} f(x \cup \{j\}), \quad x \in S
\]

(c) For the graph \((S \uparrow)\),

\[
F(x) = f(x) + \sum_{i=1}^{\infty} f(x\{i\}) = f(x) + \sum_{j \notin x} f(x \cup \{j\}), \quad x \in S
\]

Recall that the positive semigroup \((S, \subseteq)\) is a uniform; in the notation used before, if \(x \subseteq y\) then \(d(x, y) = \#(y) - \#(x)\), the length of a path from \(x\) to \(y\) in \((S, \subseteq)\). For \(n \in \mathbb{N}\), let

\[
A_n = \{x \in S : \#(x) = n\}
\]

\[
A_n(x) = \{y \in S : x \subseteq y, \#(y) = \#(x) + n\}, \quad x \in S
\]

So \(\{A_n : n \in \mathbb{N}\}\) is the partition of \(S\) induced by the function \(#\) from \(S\) to \(\mathbb{N}\), and \(\{A_n(x) : n \in \mathbb{N}\}\) is a partition of \(\{y \in S : x \subseteq y\}\). For the covering relation \(\uparrow\), note that \(x \uparrow y\) if and only if \(x \subset y\) and \(\#(y) = \#(x) + 1\). Equivalently \(y = x \cup \{i\}\) for some \(i \notin x\). More generally, let \(\uparrow^n\) denote the \(n\)-fold composition power of \(\uparrow\) for \(n \in \mathbb{N}\), so that \(x \uparrow^n y\) if and only if \(y \in A_n(x)\). (Note that \(\uparrow^0\) is the equality relation = so that \(x \uparrow^0 y\) if and only if \(y = x\).) The main point of this section is to show that if there exists a distribution with constant rate for \((S, \uparrow)\) then the distribution has constant rate for \((S, \uparrow^n)\) for each \(n \in \mathbb{N}_+\) and the distribution has constant rate for \((S, \subseteq)\).

Let \(X\) be a random variable with values in \(S\) and probability density function \(f\). Let \(F\) denote the right probability function of \(X\) for the partial order graph \((S, \subseteq)\), so that \(F(x) = \mathbb{P}(x \subseteq X)\) for \(x \in S\). For \(n \in \mathbb{N}\), let \(G_n\) denote the right probability function of \(X\) for the graph \((S, \uparrow^n)\), so that \(G_n(x) = \mathbb{P}(x \in A_n(x))\) for \(x \in S\). Of course, \(G_0 = f\). Note also that

\[
F = \sum_{n=0}^{\infty} G_n
\]

The following result gives a recursive relationship between \(G_{n+1}\) and \(G_n\) for \(n \in \mathbb{N}\).

**Theorem 16.19.** Let \(n \in \mathbb{N}\). Then

\[
G_{n+1}(x) = \frac{1}{n+1} \sum_{i \notin x} G_n(x \cup \{i\}), \quad x \in S
\]
Proof. Note that

\[ G_{n+1}(x) = \sum_{y \in A_{n+1}(x)} f(y) = \frac{1}{n+1} \sum_{i \notin x} \sum_{z \in A_n(x \cup \{i\})} f(z) = \frac{1}{n+1} \sum_{i \notin x} G_n(x \cup \{i\}), \quad x \in S \]

The factor 1/(n + 1) is present because we have counted every set y ∈ A_{n+1}(x) exactly n + 1 times in the middle double sum: we could interchange i /∈ x with any of the n + 1 elements in a set z ∈ A_n(x \cup \{i\}) without changing the set y.

**Corollary 16.15.** Suppose that X has constant rate \( \beta \in (0, \infty) \) for \((S, \uparrow^\beta)\). Then

(a) X has constant rate \( n!\beta^n \) for \((S, \uparrow^n)\).

(b) X has constant rate \( e^{-1/\beta} \) for the graph \((S, \subseteq)\).

(c) U = \#(X) has the Poisson distribution with parameter 1/\( \beta \).

**Proof.** The proofs follow easily from Theorem 16.19.

(a) We show by induction on \( n \) that

\[ G_n = \frac{1}{n!\beta^n} f \]

The result is trivially true when \( n = 0 \) and true by assumption when \( n = 1 \). Suppose that the result is true for a given \( n \in \mathbb{N}_+ \). Then By the constant rate property and by the induction hypothesis we have

\[ G_{n+1}(x) = \frac{1}{n+1} \sum_{i \notin x} G_n(x \cup \{i\}) = \frac{1}{n+1} \sum_{i \notin x} \frac{1}{n!\beta^n} f(x \cup \{i\}) \]

\[ = \frac{1}{(n+1)!\beta^n} G_1(x) = \frac{1}{(n+1)!\beta^n} \frac{1}{\beta^n} f(x) = \frac{1}{(n+1)!\beta^{n+1}} f(x), \quad x \in S \]

(b) Since X has constant rate \( n!\beta^n \) for \((S, \uparrow^n)\), we have

\[ F(x) = \sum_{n=0}^{\infty} G_n(x) = \sum_{n=0}^{\infty} \frac{1}{n!\beta^n} f(x) = e^{1/\beta} f(x), \quad x \in S \]

(c) Note that

\[ P(U = n) = P(X \in A_n) = G_n(\emptyset) = \frac{1}{n!\beta^n} f(\emptyset), \quad n \in \mathbb{N} \]

So it follows that \( f(\emptyset) = e^{-1/\beta} \) and that U has the Poisson distribution with this parameter.

But the existence question is still open. We give some examples of distributions that are not constant rate. These models are not mutually exclusive.

**Example 16.5** (Independent elements). Suppose that X is a random variable with values in \( S \) and the property that \( i \in X \) with probability \( p_i \), independently over \( i \in \mathbb{N} \). We must have \( \prod_{i \in \mathbb{N}} p_i = 0 \) so that X is finite with probability 1. For \( x \in S \),

\[ P(X = x) = \prod_{i \in x} p_i \prod_{i \in x^c} (1 - p_i) \]

while

\[ P(X \geq x) = \prod_{i \in x} p_i \]

So X has right rate function \( r \) for the graph \((S, \subseteq)\) given by \( r(x) = \prod_{i \in x^c} (1 - p_i) \) for \( x \in S \). So X has increasing rate for \((S, \subseteq)\): if \( x \subseteq y \) then \( y^c \subseteq x^c \) so

\[ r(y) = \prod_{i \in y^c} (1 - p_i) \geq \prod_{i \in x^c} (1 - p_i) = r(x) \]
Next, note that
\[ \mathbb{P}(x \uparrow X) = \sum_{j \in x^c} f(x \cup \{j\}) = \sum_{j \in x^c} \left( \prod_{i \in x} p_i \right) p_j \left( \prod_{i \in x^c} (1 - p_i) \right) \frac{1}{1 - p_j} \]
and so \( X \) has rate function \( r_1 \) for the graph \((S, \uparrow)\) given by
\[ r_1(x) = \frac{1}{\sum_{j \in x^c} p_j/(1 - p_j)}, \quad x \in S \]
Note that the denominator is the sum of the odds ratios for the events \( \{j \in X\} \), over \( j \in x^c \).

**Example 16.6** (Sampling models). Suppose that we pick a random population size \( N \in \mathbb{N} \) with probability density function \( g \). Given \( N = n \), we put \( i \in X \) independently with probability \( p \) for \( i \in \{1, \ldots, n\} \). Then for \( x \in S \),
\[ \mathbb{P}(X = x) = \sum_{n = \max(x)}^{\infty} \mathbb{P}(N = n) \mathbb{P}(X = x \mid N = n) \]
\[ = \sum_{n = \max(x)}^{\infty} g(n)p^\#(x)(1 - p)^{n - \#(x)} \]
\[ = p^\#(x) \sum_{n = \max(x)}^{\infty} g(n)(1 - p)^{n - \#(x)} \]
whereas
\[ \mathbb{P}(X \supseteq x) = \sum_{n = \max(x)}^{\infty} \mathbb{P}(N = n) \mathbb{P}(X \supseteq x \mid N = n) \]
\[ = \sum_{n = \max(x)}^{\infty} g(n)p^\#(x) = p^\#(x) \sum_{n = \max(x)}^{\infty} g(n) \]
So the rate function \( r \) of \( X \) for the graph \((S, \subseteq)\) is given by
\[ r(x) = \frac{\sum_{n = \max(x)}^{\infty} g(n)(1 - p)^{n - \#(x)}}{\sum_{n = \max(x)}^{\infty} g(n)}, \quad x \in S \]
If we generalize this model so that, given \( N = n \), \( i \in X \) with probability \( p_i \) independently for \( i \in \{1, \ldots, n\} \) then
\[ \mathbb{P}(X = x) = \prod_{i \in x} p_i \sum_{n = \max(x)}^{\infty} g(n) \prod_{i \in \{1 \ldots n\} - x} (1 - p_i) \]
whereas
\[ \mathbb{P}(X \supseteq x) = \left( \prod_{i \in x} p_i \right) \sum_{n = \max(x)}^{\infty} g(n) \]
so again, it would seem impossible for \( X \) to have constant rate for the graph \((S, \subseteq)\).

**Example 16.7** (Conditionally independent elements). Suppose that \( X \) is a random variable taking values in \( S \) with the property that for \( j \in \mathbb{N}_+ \) the events \( \{1 \in X\}, \{2 \in X\}, \ldots, \{j - 1 \in X\} \) are conditionally independent with
\[ \mathbb{P}(i \in X \mid j \in X) = \rho_{ij}, \quad i \in \{1, \ldots, j - 1\} \]
Let \( g(i) = \mathbb{P}(i \in X) \) for \( i \in \mathbb{N}_+ \). Then for \( x \in S \),
\[ \mathbb{P}(X \supseteq x) = g(\max(x)) \prod_{i \in x} p_{i, \max(x)} \]
where \( a(x) = x - \{ \max(x) \} \) and \( b(x) = \{ 1, \ldots, \max(x) - 1 \} - a(x) \). Again, it's hard to see how \( X \) could have constant rate.

**Example 16.8** (Random products of IID singletons). Suppose that \( N \) is distributed on \( \mathbb{N} \) with probability density function \( g \) so that \( g(n) = \mathbb{P}(N = n) \) for \( n \in \mathbb{N} \). We are particularly interested in the case where \( N \) has a Poisson distribution for reasons made clear in the next section. Next suppose that \( U = (U_1, U_2, \ldots) \) is an IID sequence of variables with values in \( \mathbb{N}_+ \), independent of \( N \). Let \( p_i = \mathbb{P}(U = i) \) for \( i \in \mathbb{N}_+ \) so that \( (p_i : i \in \mathbb{N}_+) \) is the common density function. Now define \( X \) with values in \( S \) by

\[
X = \{ U_1 \} \{ U_2 \} \cdots \{ U_N \}
\]

By construction, \( N = \#(X) \). As usual, let \( f \) denote the density function of \( X \) and let \( F \) denote the right probability function relative to the graph \( (S, \subseteq) \). So for \( x \in S \) with \( \#(x) = n \in \mathbb{N}_+ \), \( X = x \) if and only if \( N = n \) and \( \{ U_1 \} \{ U_2 \} \cdots \{ U_n \} \) is one of the \( n! \) factorizations of \( x \) into singletons. So in particular,

\[
f(\emptyset) = \mathbb{P}(N = 0) = g(0)
\]

For \( i \in \mathbb{N}_+ \),

\[
f(\{i\}) = g(1)\mathbb{P}(U_1 = i) = g(1)p_i
\]

For \( i, j \in \mathbb{N}_+ \) with \( i < j \),

\[
\mathbb{P}(X = \{i, j\}) = g(2)(p_ip_i + p_ip_{j-1}) = g(2)p_i(p_j + p_{j-1})
\]

In general, if \( i_1, i_2, \ldots, i_n \in \mathbb{N}_+ \) with \( i_1 < i_2 < \cdots < i_n \), then a little algebra shows that

\[
f(\{i_1, i_2, \ldots, i_n\}) = g(n)p_{i_1}(p_{i_2} + p_{i_2 - 1})(p_{i_3} + p_{i_3 - 1} + p_{i_3 - 2}) \cdots (p_{i_n} + p_{i_n - 1} + \cdots + p_{i_n - n + 1})
\]

In terms of the constant rate property, note first that \( F(\emptyset) = 1 \) so if \( X \) has constant rate then the rate constant must be \( g(0) \). Note next that \( 1 \in X \) if and only if \( N = n \) for some \( n \in \mathbb{N}_+ \) and \( U_k = 1 \) for some \( k \in \{1, 2, \ldots, n\} \). So

\[
\mathbb{P}(1 \in X) = \mathbb{P}(\{1 \subseteq X\}) = \sum_{n=0}^{\infty} g(n)[1 - (1 - p_1)^n] = (1 - p_0) - \sum_{n=1}^{\infty} g(n)(1 - p_1)^n
\]

In this case that \( N \) has the Poisson distribution with parameter \( \lambda \in (0, \infty) \) this reduces to

\[
\mathbb{P}(1 \in X) = (1 - e^{-\lambda}) - e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{n!} (1 - p_1)^n
\]

\[
= (1 - e^{-\lambda}) - e^{-\lambda}[e^{\lambda(p_1 - 1)} - 1] = 1 - e^{\lambda p_1}
\]

So if \( X \) has constant rate (which must be \( e^{-\lambda} \)) then we must have

\[
\lambda e^{-\lambda} p_1 = e^{-\lambda}(1 - e^{-\lambda} p_1)
\]

or equivalently \( \lambda p_1 + e^{-\lambda} p_1 = 1 \), which is impossible since \( \lambda \in (0, \infty) \) and \( p_1 \in (0, 1) \).

Returning to the general case, suppose that \( X \) has probability density function \( f \) and right probability function \( F \) (relative to \( (S, \subseteq) \)). Recall that the right probability function \( F \) determines the density function \( f \):

\[
f(x) = \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_n(x)} F(y), \quad x \in S
\]

Thus, if \( X \) has constant rate \( \alpha \) then

\[
f(x) = \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_n(x)} f(y), \quad x \in S
\]  

(16.15)

We showed above that if \( X \) has constant rate for \( (S, \uparrow) \) then \( X \) also has constant rate for \( (S, \subseteq) \). We do now know if the converse is true. However, we also know that if \( X \) has constant rate for \( (S, \uparrow) \) then \( U = \#(X) \) has a Poisson distribution. Our next discussion will show that this is also true if \( X \) has constant rate for \( (S, \subseteq) \).
Theorem 16.20. Suppose that \( X \) has constant rate \( \alpha \in (0, 1) \). Then
\[
\alpha \mathbb{P}(U = k) = \mathbb{E}
\left[
(-1)^{U+k} \binom{U}{k}
\right], \quad k \in \mathbb{N}
\] (16.16)

Hence \( U \) has the Poisson distribution with parameter \(- \ln \alpha\).

**Proof.** Let \( f \) denote the probability density function of \( X \), as above. From (16.15),
\[
\mathbb{P}(U = k) = \sum_{x \in A_k} f(x) = \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \sum_{x \in A_k} \sum_{y \in A_n(x)} f(y)
\]
The last two sums are over all \( x, y \in S \) with \( \#(y) = n + k \), and \( x \subseteq y \). Interchanging the order of summation gives
\[
\mathbb{P}(U = k) = \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_{n+k}} \sum \{ f(y) : x \in A_k, \; x \subseteq y \}
\]
\[
= \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_{n+k}} \binom{n+k}{k} f(y)
\]
\[
= \frac{1}{\alpha} \sum_{n=0}^{\infty} (-1)^n \binom{n+k}{k} \mathbb{P}(U = n + k)
\]
Equivalently (with the substitution \( j = n + k \)),
\[
\alpha \mathbb{P}(U = k) = \sum_{j=k}^{\infty} (-1)^{j-k} \binom{j}{k} \mathbb{P}(U = j)
\]

With the usual convention on binomial coefficients, that is \( \binom{a}{b} = 0 \) if \( b < 0 \) or \( b > a \), we have
\[
\alpha \mathbb{P}(U = k) = \sum_{j=0}^{\infty} (-1)^{j+k} \binom{j}{k} \mathbb{P}(U = j) = \mathbb{E}
\left[
(-1)^{k+U} \binom{U}{k}
\right]
\]
The fact that \( U \) has a Poisson distribution with parameter \(- \ln \alpha\) now follows from the characterizations in Appendix A. \( \square \)

Our last discussion concerns the existence of a constant rate distribution. Let \( g_k \) denote the conditional probability density function of \( X \) given that \( U = k \), so that
\[
g_k(x) = \mathbb{P}(X = x \mid U = k), \quad x \in A_k
\]
and suppose that \( U \) has the Poisson distribution with parameter \( \lambda \). Then the condition for \( X \) to have constant rate \( e^{-\lambda} \) is
\[
e^{-2\lambda} \frac{\lambda^k}{k!} g_k(x) = \sum_{n=0}^{\infty} (-1)^n \sum_{y \in A_n(x)} e^{-\lambda} \frac{\lambda^{n+k}}{(n+k)!} g_{n+k}(y), \quad x \in A_k
\]
or equivalently,
\[
e^{-\lambda} \frac{\lambda^k}{k!} g_k(x) = \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{(n+k)!} \sum_{y \in A_n(x)} g_{n+k}(y), \quad x \in A_k
\] (16.17)

**Lemma 16.1.** The previous condition holds if
\[
\sum_{y \in A_n(x)} g_{n+k}(y) = \binom{n+k}{k} g_k(x), \quad x \in A_k, \; k, n \in \mathbb{N}
\] (16.18)
CHAPTER 16. THE SUBSET SEMIGROUP

174

Proof. Suppose that the previous condition holds. Then for \( x \in A_k \),

\[
\sum_{n=0}^{\infty} \frac{(-\lambda)^n}{(n+k)!} \sum_{y \in A_n(x)} g_{n+k}(y) = \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{(n+k)!} \frac{(n+k)!}{n!k!} g_k(x)
\]

\[
= \frac{g_k(x)}{k!} \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} = e^{-\lambda} g_k(x)
\]

and thus (1) holds.

The last condition is equivalent to

\[
g_k(x) = \frac{1}{\binom{n}{k}} \sum_{y \in B_n(x)} g_n(y), \quad k \in \mathbb{N}, n \in \mathbb{N}, k \leq n, x \in B_k
\]  

(16.19)

where \( B_k = \{ x \in S : \#(x) = k \} \) and \( B_n(x) = \{ y \in S : \#(y) = n, x \subseteq y \} \). Condition (1) has the following interpretation:

Theorem 16.21. Suppose that \( Y \) is a random variable taking values in \( B_n \) with probability density function \( g_n \). Let \( X \) be a randomly chosen subset of \( Y \) of size \( k \leq n \). Then the density function \( g_k \) of \( X \) satisfies (1).

Proof. The precise meaning of the hypothesis is that given \( Y = y \) where \( y \in B_n \), the conditional distribution of \( X \) is uniform on the collection of subsets of \( y \) of size \( k \). Thus, for \( x \in B_k \),

\[
g_k(x) = \mathbb{P}(X = x) = \sum_{y \in B_n(x)} \mathbb{P}(y = y) \mathbb{P}(X = x \mid Y = y) = \frac{1}{\binom{n}{k}} \sum_{y \in B_n(x)} g_n(y)
\]

\[\square\]

Note 16.1. In particular, the last condition is consistent. If \( g_n \) is a probability density function on \( B_n \), for a given \( n \in \mathbb{N} \), then we could define \( g_k \) by the last condition for \( k \in \{0, \ldots, n-1\} \); these would also be probability density functions.

Theorem 16.22. No sequence of probability density functions \( (g_n : n \in \mathbb{N}) \) satisfying the last condition exists.

Proof. Suppose that such a sequence does exist. Fix \( k \in \mathbb{N} \) and \( x \in B_k \) and then let \( n \to \infty \) in the last condition. Since

\[
\sum_{y \in B_n(x)} g_n(y) \leq 1
\]

we must have \( g_k(x) = 0 \).

\[\square\]
Chapter 17

Induced Graphs

The various models in this chapter are a good source of counterexamples and are good illustrations of some of the general theory. For the basic setup, we suppose that $(S, \mathcal{F}, \lambda)$ is a measure space and that $S$ has a measurable partition $\mathcal{P} = \{S_n : n \in \mathbb{N}\}$ with $a_n = \lambda(S_n) \in (0, \infty)$ for each $n \in \mathbb{N}$. Define the index function $\varphi : S \to \mathbb{N}$ by $\varphi(x) = n$ if $x \in S_n$. The assumption that $\lambda(S_n) < \infty$ for each $n \in \mathbb{N}$ is necessary because otherwise our graphs of interest would not be locally finite. Local finiteness is an essential part of the general theory, particularly for the existence of constant rate distributions. In the discrete case, $S$ is countable and $a_n = \#(S_n) \in \mathbb{N}_+$ is for each $n \in \mathbb{N}$. In this chapter, functions on $\mathbb{N}$ are usually given in subscript form.

In the first three sections, we will consider graphs partitioned, in the sense of Chapter 11, by the standard graphs on $\mathbb{N}$ studied in Chapter 11. After that, in the discrete case, we will consider the lexicographic sum $(S, \preceq)$ of $(S, =)$ over $(\mathbb{N}, <)$ in the sense of Chapter 8.4. The final two sections consider a special case in which $(S, \preceq)$ corresponds to a positive semigroup.

Throughout this chapter, $X$ denotes a random variable with values in $S$ and with density function $f$ with respect to $\lambda$. Let $N = \varphi(X)$ denote the corresponding index variable with values in $\mathbb{N}$, so that by $N = n$ if and only if $X \in S_n$ for $n \in \mathbb{N}$. Let $p$ denote the discrete density of $N$ so that

$$p_n = P(N = n) = P(X \in S_n) = \int_{S_n} f(x) d\lambda(x), \quad n \in \mathbb{N}$$

For the first three sections, recall that if $\Rightarrow$ is a relation on $\mathbb{N}$ then the induced relation $\Rightarrow$ on $S$ is defined by $x \Rightarrow y$ if and only if $\varphi(x) \to \varphi(y)$ for $x$, $y \in S$. That is, $x \Rightarrow y$ if and only if $x \in S_m$ and $y \in S_n$ for some $m, n \in \mathbb{N}$ with $m \to n$. We then let $F$ denote the right probability function of $X$ with respect to the graph $(S, \Rightarrow)$ and $P$ the right probability function of $N$ with respect to $(\mathbb{N}, \to)$ so that

$$F(x) = P(x \Rightarrow X) = P(n \Rightarrow N) = P_n, \quad x \in S_n, n \in \mathbb{N}$$

17.1 The Cover Graph

Let $\uparrow$ denote the relation on $\mathbb{N}$ given by $n \uparrow n + 1$ for each $n \in \mathbb{N}$. That is, $(\mathbb{N}, \uparrow)$ is the covering graph for the standard discrete total order graph $(\mathbb{N}, \leq)$. The corresponding partitioned graph is $(S, \uparrow)$ so that $x \uparrow y$ if and only if $x \in S_n$ and $y \in S_{n+1}$ for some $n \in \mathbb{N}$.

Exercise 17.1. In the discrete case, sketch the graph $(S, \uparrow)$ in each of the following cases:

(a) $a_n = 2$ for $n \in \mathbb{N}$.
(b) $a_n = n + 1$ for $n \in \mathbb{N}$.

Our first goal is to compute the left walk functions for $(S, \uparrow)$. As with all partitioned graphs, the walk functions are constant on the partition sets $S_n$ for $n \in \mathbb{N}$.

Proposition 17.1. The left walk function $\gamma_m$ of order $m \in \mathbb{N}$ for $(S, \uparrow)$ is given by

$$\gamma_m(x) = a_{n-m}a_{n-m+1} \cdots a_{n-1}, \quad x \in S_n, n \in \{m, m+1, \ldots\}$$

and $\gamma_m(x) = 0$ otherwise.
Theorem 17.1. A distribution on $S$ with constant rate $\alpha \in (0, \infty)$ for $(S, \uparrow)$ exists if and only if
\[
\frac{1}{p_0} = \sum_{n=0}^{\infty} \frac{1}{\alpha^n a_0 a_1 \cdots a_{n-1}} < \infty.
\] (17.1)

In this case, the probability density function $f$ is given by
\[
f(x) = \frac{1}{\alpha^n a_0 a_1 \cdots a_n} p_0, \quad x \in S_n, \ n \in \mathbb{N}
\]

Proof. The constant rate condition is $f(x) = \alpha F(x)$ for $x \in S$, so $f(x) = \alpha p_{n+1}$ for $x \in S_n$ and $n \in \mathbb{N}$. Integrating over $S_n$ gives $p_n = \alpha a_n p_{n+1}$ for $n \in \mathbb{N}$, or equivalently
\[
p_{n+1} = \frac{1}{\alpha a_n} p_n, \quad n \in \mathbb{N}
\]

Solving gives
\[
p_n = \frac{1}{\alpha^n a_0 a_1 \cdots a_{n-1}} p_0, \quad n \in \mathbb{N}
\]

The condition $\sum_{n=0}^{\infty} p_n = 1$ then requires that
\[
\frac{1}{p_0} = \sum_{n=0}^{\infty} \frac{1}{\alpha^n a_0 a_1 \cdots a_{n-1}}
\]

Finally, substituting into $f(x) = \alpha p_{n+1}$ for $x \in S_n$ and $n \in \mathbb{N}$ gives the result. \qed

The proof of Theorem 17.1 also follows from basic results in Chapter ?? on partitioned graphs. In particular, a constant rate distribution exists if $a_n$ is bounded away from 0 for $n \in \mathbb{N}$.

Corollary 17.1. If $a_n \geq a$ for all $n \in \mathbb{N}$ where $a \in (0, \infty)$ then a distribution with constant rate $\alpha$ for $(S, \uparrow)$ exists for all $\alpha \in (1/a, \infty)$

Proof. Since $a_n \geq a > 0$ for all $n \in \mathbb{N}$ we have
\[
\sum_{n=0}^{\infty} \frac{1}{\alpha^n a_0 \cdots a_n} \leq \sum_{n=0}^{\infty} \left( \frac{1}{\alpha a} \right)^n
\]

The sum converges if $\alpha a > 1$. \qed

In the discrete case, $a_n \in \mathbb{N}_+$ for each $n \in \mathbb{N}$ and hence a constant rate distribution exists for every $\alpha \in (1, \infty)$.

Corollary 17.2. Suppose that $X$ has constant rate $\alpha \in (1, \infty)$ for $(S, \uparrow)$. Then
(a) $N$ has probability density function $p$ given by
\[ p_n = \frac{1}{\alpha^n a_0 a_1 \cdots a_{n-1}} p_0, \quad n \in \mathbb{N} \]

(b) $N$ has rate function $r$ for $(\mathbb{N}, \uparrow)$ given by
\[ r_n = \alpha a_n, \quad n \in \mathbb{N} \]

(c) For $n \in \mathbb{N}$, the conditional distribution of $X$ given $N = n$ is uniform on $S_n$.

(*Proof. *) Part (a) follows directly from the proof of Theorem 17.1 and (b) follows from (a). Part (c) is clear since the density function $f$ of $X$ is constant on $S_n$ for each $n \in \mathbb{N}$.

So for the graph $(\mathbb{N}, \uparrow)$, $N$ has increasing, decreasing, or constant rate if $a_n$ is increasing, decreasing, or constant in $n \in \mathbb{N}$, respectively. In the discrete case, note that if $a_n \in \mathbb{N}_+$ is decreasing in $n \in \mathbb{N}$ then $a_n$ is eventually constant in $n \in \mathbb{N}$. The distributions defined in part (a) are interesting, and include geometric and Poisson distributions as special cases.

**Example 17.1.** Suppose that $a_n = a \in (0, \infty)$ for all $n \in \mathbb{N}$ and that $\alpha > 1/a$. Then $p_0 = (aa - 1)/aa$ so $X$ has density function $f$ given by
\[ f(x) = \frac{aa - 1}{(aa)^{n+1}}, \quad x \in S_n, n \in \mathbb{N} \]
and $N$ has density function $p$ given by
\[ p_n = \frac{aa - 1}{(aa)^{n+1}} = \frac{aa - 1}{aa} \left( \frac{1}{aa} \right)^n, \quad n \in \mathbb{N} \]

That is, $N$ has the geometric distribution on $\mathbb{N}$ with success parameter $(aa - 1)/aa$. As shown in Chapter 11, the geometric distribution has constant rate for each of the graphs $(\mathbb{N}, \leq)$, $(\mathbb{N}, <)$, and $(\mathbb{N}, \uparrow)$.

**Example 17.2.** Suppose that $a_n = n + 1$ for $n \in \mathbb{N}$ and that $\alpha > 0$. Then $p_0 = e^{-1/\alpha}$ so $X$ has density function $f$ given by
\[ f(x) = e^{-1/\alpha} \frac{1}{\alpha^n (n+1)!}, \quad x \in S_n, n \in \mathbb{N} \]
and $N$ has density function $p$ given by
\[ p_n = e^{-1/\alpha} \frac{1}{\alpha^n n!}, \quad n \in \mathbb{N} \]

That is, $N$ has the Poisson distribution with parameter $1/\alpha$, and has increasing rate for $(\mathbb{N}, \uparrow)$.

**Exercise 17.3.** In the discrete case, find results analogous to Theorem 17.1, Corollary 17.1, and Corollary 17.2 for the reflexive completion of $(S, \uparrow)$.

Finally, suppose that (17.1) holds, so that the density function $f$ with constant rate $\alpha \in (0, \infty)$ exists, and consider the random walk $Y = (Y_1, Y_2, \ldots)$ on $(S, \uparrow)$ associated with $f$. So in particular, $Y_1$ has density function $f$ and $Y_m \uparrow Y_{m+1}$ for each $m \in \mathbb{N}_+$.

**Theorem 17.2.** For $m \in \mathbb{N}_+$, $Y_m$ has probability density function $f_m$ given by
\[ f_m(x) = \frac{1}{\alpha^{m+1} a_0 \cdots a_{m-1} a_m} p_0, \quad x \in S_n, n \in \{m - 1, m, m + 1, \ldots\} \]

(*Proof. *) From our general theory, $Y_m$ has density function $f_m$ given by
\[ f_m(x) = \gamma_{m-1}(x) \alpha^n F(x) = \gamma_{m-1}(x) \alpha^{n-1} f(x), \quad x \in S \]
where $\gamma_{m-1}$ is the left walk function of order $m-1$ and $F$ the right probability function for $(S, \uparrow)$. Substituting the results from Proposition 17.1 and Theorem ?? gives the result. 

Note that $f_m(x) = 0$ for $x \in S_n$ with $n < m - 1$. 

\[ \square \]
Corollary 17.3. Let $N_m = \varphi(Y_m)$, the index variable of $Y_m$, for $m \in \mathbb{N}_+$. Then

(a) $N_m = N_1 + (m - 1)$ for $m \in \mathbb{N}_+$.

(b) The density function of $N_m$ is given by

$$P(N_m = n) = \frac{1}{\alpha_n - 1} g(n), \quad n \in \{m - 1, m, m + 1, \ldots\}$$

(c) For $n \in \{m - 1, m, m + 1\}$, the conditional distribution of $Y_m$ given $N_m = n$ is uniform on $S_n$.

Proof. The proofs are straightforward.

(a) Since $Y_m \uparrow Y_{m+1}$, it follows that $N_{m+1} = N_m + 1$ for $m \in \mathbb{N}_+$.

(b) This follows from (a)

$$P(N_m = n) = P(N_1 = n - m + 1) = p_{n-m+1}, \quad n \in \{m - 1, m, m + 1, \ldots\}$$

(c) As before, this follows since $f_m$ is constant on $S_n$ for $n \in \{m - 1, m, m + 1, \ldots\}$.

We know from the general theory that the constant rate distribution governs the most random way to put points in the graph $(S, \uparrow)$, in the sense that given $Y_{m+1} = x \in S$, the random sequence $(Y_1, Y_2, \ldots, Y_m)$ is uniformly distributed on

$$\{(x_1, x_2, \ldots, x_m) \in S^m : x_1 \uparrow x_2 \uparrow \cdots \uparrow x_m \uparrow x\}$$

For this simple graph, we can say more: given $Y_m = x \in S_n$ for $n \in \{m - 1, m, m + 1, \ldots\}$, $Y_{m-1}$ is uniformly distribution on $S_{n-1}$, $Y_{m-2}$ is uniformly distributed on $S_{n-2}$, and so forth, with $Y_1$ uniformly distributed on $S_{n-m+1}$.

Exercise 17.4. Find the right rate function of $Y_m$ relative to $(S, \uparrow)$ for $m \in \mathbb{N}_+$.

17.2 The Completed Cover Graph

Let $\rightarrow$ denote the relation on $\mathbb{N}$ given by $n \rightarrow n$ and $n \rightarrow n + 1$ for $n \in \mathbb{N}$, so that $(\mathbb{N}, \rightarrow)$ is the reflexive completion of the graph $(\mathbb{N}, \uparrow)$ in the last section. The graph $(S, \Rightarrow)$ partitioned by $(\mathbb{N}, \rightarrow)$ is given by $x \Rightarrow y$ if and only if for some $n \in \mathbb{N}$, either $x, y \in S_n$ or $x \in S_n$ and $y \in S_{n+1}$. Note that $(S, \Rightarrow)$ is not the reflexive completion of the graph $(S, \uparrow)$ studied in the last section.

Hence $G(n) = g(n) + g(n + 1)$ for $n \in \mathbb{N}$. The basic equation in (??) for a constant rate distribution on $(S, \Rightarrow)$ is $g(n) = \alpha a_n[g(n) + g(n + 1)]$ for $n \in \mathbb{N}$ or equivalently,

$$g(n + 1) = \left(\frac{1}{\alpha a_n} - 1\right) g(n), \quad n \in \mathbb{N}$$

Solving gives

$$g(n) = g(0) \prod_{k=0}^{n-1} \left(\frac{1}{\alpha a_k} - 1\right), \quad n \in \mathbb{N}$$

Hence a constant rate distribution exists if and only if $\alpha a_n < 1$ for all $n \in \mathbb{N}$ and

$$c(\alpha) := \sum_{n=0}^{\infty} \prod_{k=0}^{n-1} \left(\frac{1}{\alpha a_k} - 1\right) < \infty$$

in which case the discrete density function on $\mathbb{N}$ is given by

$$g(n) = \frac{1}{c(\alpha)} \prod_{k=0}^{n-1} \left(\frac{1}{\alpha a_k} - 1\right), \quad G(n) = \frac{1}{c(\alpha)} \prod_{k=0}^{n-1} \left(\frac{1}{\alpha a_k} - 1\right) \frac{1}{\alpha a_n}; \quad n \in \mathbb{N}$$
In particular, a constant rate distribution for \((S, \Rightarrow)\) given by

\[
f(x) = \frac{1}{c(\alpha)} \left[ \prod_{k=0}^{n-1} \left( \frac{1}{\alpha a_k} - 1 \right) \right] \frac{1}{a_n}, \quad x \in S_n, \; n \in \mathbb{N}
\]

In particular, if \(a_n\) is bounded away from 0 and \(\infty\) in \(n \in \mathbb{N}\), in a certain way, then a constant rate distribution exists.

**Theorem 17.3.** Suppose that \(a \leq a_n \leq b\) for all \(n \in \mathbb{N}\) where \(0 < a < b < 2a < \infty\). Then a distribution with constant rate \(\alpha\) for the graph \((S, \Rightarrow)\) partitioned by \((N, \gamma)\) exists for all \(\alpha \in \left( \frac{b}{2}, \frac{1}{a} \right)\).

**Proof.** First, \(\alpha a_n < \frac{1}{2}b = 1\) for all \(n \in \mathbb{N}\). Next, \(\alpha > \frac{1}{2a}\) so \(\alpha \geq \frac{1}{(1+r)a}\) for some \(r \in (0, 1)\). Hence \(\alpha a_n \geq \frac{1}{(1+r)a} = \frac{1}{1+r}\) and so \(\frac{1}{\alpha a_n} - 1 \leq r\). Therefore

\[
\sum_{n=0}^{\infty} \prod_{k=0}^{n-1} \left( \frac{1}{\alpha a_k} - 1 \right) \leq \sum_{k=0}^{\infty} r^n < \infty
\]

\[\square\]

### 17.3 The Strict Partial Order Graph

Next we consider the graph \((S, \prec)\) partitioned by \((N, \prec)\). So \(x \prec y\) if and only if \(x \in S_m\) and \(y \in S_n\) for some \(m, n \in \mathbb{N}\) with \(m < n\). As the notation suggests, \(\prec\) is a strict partial order, since clearly \(\prec\) is anti-reflexive, anti-symmetric, and transitive. As usual, we start with the left walk functions.

**Proposition 17.2.** The left walk function \(\gamma_k\) of order \(k \in \mathbb{N}_+\) is given by

\[
\gamma_k(x) = \sum_{n_1 < n_2 < \cdots < n_k < n} a_{n_1} a_{n_2} \cdots a_{n_k}, \quad x \in S_n, \; n \in \mathbb{N}
\]

**Proof.** This follows directly from the general theory of partitioned graphs. \[\square\]

So the left walk function does not have a simple closed form, except in special cases.

**Example 17.3.** Suppose that \(a_n = a \in (0, \infty)\) for each \(n \in \mathbb{N}\). Then for \(k \in \mathbb{N}_+\),

\[
\gamma_k(x) = \binom{n}{k} a^k, \quad x \in S_n, \; n \in \mathbb{N}
\]

As before, suppose that \(X\) is a random variable with values in \(S\) and probability density function \(f\), and with right probability function \(F\) for \((S, \ulcorner)\). Let \(N = \phi(X)\) denote the index function of \(X\) with values in \(\mathbb{N}\) and density function \(p\), and with right probability function \(P\) for \((N, \prec)\).

The discrete right probability function \(G\) is related to the discrete density \(g\) by \(G(n) = \sum_{k=n+1}^{\infty} g(k)\).

Equivalently \(g(0) = 1 - G(0)\) and \(g(n) = G(n-1) - G(n)\) for \(n \in \mathbb{N}_+\). Again, the basic equation in (?) become \(g(n) = \alpha a_n G(n)\) for \(n \in \mathbb{N}\). Solving we have

\[
G(n) = \frac{1}{(1 + \alpha a_0) \cdots (1 + \alpha a_n)}, \quad n \in \mathbb{N}
\]

If \(G(n) \to 0\) as \(n \to \infty\) then \(G\) is a valid right probability function and the corresponding discrete probability density function \(g\) is given by

\[
g(n) = \frac{\alpha a_n}{(1 + \alpha a_0) \cdots (1 + \alpha a_n)}, \quad n \in \mathbb{N}
\]

Then the distribution with constant rate \(\alpha\) for \((S, \Rightarrow)\) has right probability function \(F\) and density \(f\) given by

\[
F(x) = \frac{1}{(1 + \alpha a_0) \cdots (1 + \alpha a_n)}, \quad f(x) = \alpha \frac{1}{(1 + \alpha a_0) \cdots (1 + \alpha a_n)}; \quad x \in S_n, \; n \in \mathbb{N}
\]

In particular, a constant rate distribution for \((S, \Rightarrow)\) exists if \(a_n\) is bounded away from 0 in \(n \in \mathbb{N}\).
Theorem 17.4. If \( a_n \geq a \) for all \( n \in \mathbb{N} \) where \( a \in (0, \infty) \) then a constant rate distribution for for the graph \((S, \Rightarrow)\) partitioned by \((\mathbb{N}, \prec)\) exists for all \( \alpha \in (0, \infty) \).

Proof. If \( a_n \geq a > 0 \) for all \( n \in \mathbb{N} \) then

\[
G(n) \leq \frac{1}{(1 + \alpha a)^n} \to 0 \text{ as } n \to \infty
\]

\[
\square
\]

17.4 The Partial Order Graph

Suppose first that \((S, \Rightarrow)\) is a graph partitioned by \((\mathbb{N}, \leq)\), so \( x \Rightarrow y \) if and only if \( x \in S_m \) and \( y \in S_n \) with \( m \leq n \). The basic equations in (8.5) for the existence of a distribution with constant rate \( \alpha \in (0, \infty) \) for \((S, \Rightarrow)\) are

\[
g(n) = \alpha a_n G(n), \quad n \in \mathbb{N}
\]

Solving gives

\[
G(n) = \prod_{k=0}^{n-1} (1 - \alpha a_k), \quad g(n) = \alpha a_n \prod_{k=0}^{n-1} (1 - \alpha a_k); \quad n \in \mathbb{N}
\]

assuming that \( \alpha \) satisfies \( 0 < 1 - \alpha a_k < 1 \) for each \( k \in \mathbb{N} \) and \( \prod_{k=0}^{\infty} (1 - \alpha a_k) = 0 \). Then, the distribution with constant rate \( \alpha \) for \((S, \Rightarrow)\) has density function \( f \) and right probability function \( F \) given by

\[
f(x) = \alpha \prod_{k=0}^{n-1} (1 - \alpha a_k), \quad F(x) = \prod_{k=0}^{n-1} (1 - \alpha a_k); \quad x \in S_n, \ n \in \mathbb{N}
\]

In particular, if \( a_n \) is bounded away from 0 and \( \infty \) in \( n \in \mathbb{N} \), then there exist constant rate distributions.

Theorem 17.5. Suppose that \( a \leq a_n \leq b \) for \( n \in \mathbb{N} \) where \( 0 < a < b < \infty \). Then there a distribution with constant rate \( \alpha \) for the graph \((S, \Rightarrow)\) partitioned by \((\mathbb{N}, <)\) exists for each \( \alpha \in (0, 1/b) \).

Proof. Since \( a \leq a_n \leq b \) for \( n \in \mathbb{N} \) we have \( 1 - a b \leq 1 - a a_n \leq 1 - a a \) for \( n \in \mathbb{N} \). But \( 0 < a < 1/b < 1/a \) so \( 1 - a b > 0 \) and \( 1 - a a < 1 \).

\[
\square
\]

17.5 The Lexicographic Sum

In this section, we assume that our underlying space is discrete, so that \( S \) is countable and the reference measure \( # \) is counting measure. Define the partial order \( \preceq \) on \( S \) by \( x \prec y \) if and only if \( \phi(x) < \phi(y) \). That is, for \((x, y) \in S^2, x \preceq y \) if and only if \( x = y \), or \( x \in S_m, y \in S_n \) for some \( m, n \in \mathbb{N} \) with \( m < n \). The poset \((S, \preceq)\) can be constructed as the lexicographic sum of the graphs \((S_n, =)\) over \((\mathbb{N}, <)\) as described in Section 8.4. The elements in \( S_0 \) are minimal, and in particular, if \( S_0 = \{e\} \) then \( e \) is the minimum element of \( S \). Note that the covering graph of \((S, \preceq)\) is the graph \((S, \uparrow)\) partitioned by \((\mathbb{N}, \uparrow)\) studied above. A special case in which \((S, \preceq)\) is associated with a positive semigroup will be studied in the next section.

The walk function for the partial order graph \((S, \preceq)\) is considerably more complicated that for our previous graphs.

Proposition 17.3. The left walk function \( \gamma_j \) of order \( j \in \mathbb{N} \) for \((S, \preceq)\) is given by

\[
\gamma_j(x) = \sum_{l=0}^{n} \binom{j}{l} \sum \left\{ \prod_{i \in J} k_i : J \subseteq \{0, 1, \ldots, n - 1\}, \ #(J) = l \right\}, \quad x \in S_n, \ n \in \mathbb{N}
\]

Proof. For the proof, it helps to have some additional notation. If \( J \subseteq \mathbb{N} \) is finite, define \( k_J = \prod_{j \in J} k_j \). For \( n \in \mathbb{N} \) and \( l \in \{0, 1, \ldots, n\} \) define

\[
J_{n, l} = \{ J \subseteq \{0, 1, \ldots, n - 1\} : \ #(J) = l \}
\]
Note that $J_{n,0} = \{\emptyset\}$. Recall that $\gamma_0(x) = 1$ for $x \in S$ and $\gamma_j+1(x) = \sum_{y \leq x} \gamma_j(y)$ for $x \in S$ and $j \in \mathbb{N}$. So we can give a proof by induction on $j$. The formula is vacuously true when $j = 0$. Suppose now that the formula holds for a given $j \in \mathbb{N}$ and $x \in S$. Then for $x \in S_n$,

$$
\gamma_{j+1}(x) = \sum_{y \leq x} \gamma_j(y) = \gamma_j(x) + \sum_{m=0}^{n-1} \sum_{y \in S_m} \gamma_j(y)
$$

$$
= \sum_{l=0}^{n} \left( \begin{array}{c} j \\ l \end{array} \right) \sum_{J \in \mathcal{J}_n,l} k_j + \sum_{m=0}^{n-1} k_m \sum_{l=0}^{j} \left( \begin{array}{c} j \\ l \end{array} \right) \sum_{J \in \mathcal{J}_{m,l}} k_j
$$

$$
= \sum_{l=0}^{n} \left( \begin{array}{c} j \\ l \end{array} \right) \sum_{J \in \mathcal{J}_n,l} k_j + \sum_{l=0}^{j} \left( \begin{array}{c} j \\ l \end{array} \right) \sum_{J \in \mathcal{J}_n,l} k_j
$$

$$
= \sum_{l=0}^{n} \left( \begin{array}{c} j + 1 \\ l \end{array} \right) \sum_{J \in \mathcal{J}_n,l} k_j
$$

Exercise 17.5. Suppose that $x \in S_3$. Explicitly compute $\gamma_j(x)$ for $j \in \{0, 1, 2, 3, 4\}$.

Next we compute the Möbius function $M$ for the partial order graph $(S, \preceq)$. Since $M(x, y) = 1$ for $x \in S$ and $M(x, y) = 0$ if $x \npreceq y$, the only values of interest are $M(x, y)$ with $x \in S_m$ and $y \in S_{m+n}$ with $m \in \mathbb{N}$ and $n \in \mathbb{N}_+$. As with the walk function, it’s not surprising that $M$ is constant for such $(x, y)$.

Proposition 17.4. The Möbius function $M$ for $(S, \preceq)$ is given as follows, for $m \in \mathbb{N}$ and $n \in \mathbb{N}_+$:

$$
M(x, y) = \sum_{l=0}^{n-1} (-1)^{l+1} \prod_{J \in J} k_{m+j} : J \subseteq \{1, \ldots, n-1\}, \#(J) = l, \quad x \in S_m, y \in S_{m+n}
$$

Proof. Suppose that $m \in \mathbb{N}$ and that $x \in S_m$. If $y \in S_{m+1}$ then

$$
M(x, y) = - \sum_{t \in [x, y]} M(x, t) = -M(x, x) = -1
$$

Similarly, if $y \in S_{m+2}$ then

$$
M(x, y) = - \sum_{t \in [x, y]} M(x, t) = - \left[ M(x, x) + \sum_{t \in S_{m+1}} M(x, t) \right] = -[1 - k_{m+1}] = -1 + k_{m+1}
$$

For $y \in S_{m+3}$,

$$
M(x, y) = - \sum_{t \in [x, y]} M(x, t) = - \left[ M(x, x) + \sum_{t \in S_{m+1}} M(x, t) + \sum_{t \in S_{m+2}} M(x, t) \right]
$$

$$
= -[1 - k_{m+1} - k_{m+2}(-1 + k_{m+1})] = -1 + k_{m+1} + k_{m+2} - k_{m+1} k_{m+2}
$$

Continuing in this way gives the result. A more formal induction proof over $n$ is as follows. For $n \in \mathbb{N}_+$ and $j \in \{0, 1, \ldots, n-1\}$, let $\mathcal{J}_{n,l} = \{ J \subseteq \{1, 2, \ldots, n-1\} : \#(J) = l \}$. As shown above, the formula is true
Suppose now that $n = 1$. Assume the formula holds for $m \in \mathbb{N}$ and for positive integers $r \leq n$ for a given $n \in \mathbb{N}_+$. Let $x \in S_m$ and $y \in S_{m+n+1}$. Then

$$M(x, y) = - \sum_{t \in [x,y]} M(x, t) = - \left[ M(x, x) + \sum_{r=1}^{n} \sum_{t \in S_{m+r}} M(x, t) \right]$$

$$= - \left[ 1 + \sum_{r=1}^{n} k_{m+r} \sum_{l=0}^{r-1} (-1)^{l+1} \sum \left\{ \prod_{j \in J} k_{m+j} : J \in J_{r,l} \right\} \right]$$

$$= - \left[ 1 + \sum_{l=0}^{n-1} (-1)^{l+1} \sum_{r=l+1}^{n} \sum \left\{ \prod_{j \in J} k_{m+j} : J \in J_{n+1,l} \right\} \right]$$

$$= - \left[ 1 + \sum_{l=0}^{n} (-1)^{l+1} \sum \left\{ \prod_{j \in J} k_{m+j} : J \in J_{n+1,l} \right\} \right]$$

$$= \sum_{l=0}^{n} (-1)^{l+1} \sum \left\{ \prod_{j \in J} k_{m+j} : J \in J_{n+1,l} \right\}$$

$\square$

### 17.6 Distributions

Suppose now that $X$ is a random variable with values in $S$ and with probability density function $f$. Let $N = \phi(X)$ denote the corresponding index random variable with values in $\mathbb{N}$, so that $N = n$ if and only if $X \in S_n$ for $n \in \mathbb{N}$. Thus $N$ has density function $(p_n : n \in \mathbb{N})$ given by

$$p_n = \mathbb{P}(N = n) = \mathbb{P}(X \in S_n) = \sum_{x \in S_n} f(x), \quad n \in \mathbb{N}$$

Let $F$ and $G$ denote the right probability functions of $X$ for the graphs $(S, \leq)$ and $(S, \uparrow)$ respectively. Then

$$F(x) = f(x) + \sum_{m=n+1}^{\infty} p_m, \quad n \in \mathbb{N}, x \in S_n$$

On the other hand, $f$ can be recovered from $F$ via M"obius inversion under certain conditions. For $m \in \mathbb{N}$ and $n \in \mathbb{N}_+$, let $M_{m,m+n}$ denote the constant value of the M"obius function $M(x, y)$ given in Proposition 17.4 when $x \in S_m$ and $y \in S_{m+n}$.

**Proposition 17.5.** Assuming absolute convergence of the series,

$$f(x) = F(x) + \sum_{n=1}^{\infty} M_{m,m+n} \sum_{y \in S_{m+n}} F(y), \quad m \in \mathbb{N}, x \in S_m$$

**Proof.** Let $m \in \mathbb{N}$ and $x \in S_m$. By the M"obius inversion formula, again assuming absolute convergence of the series,

$$f(x) = \sum_{x \leq y} M(x, y)F(y) = F(x) + \sum_{x < y} M(x, y)F(y)$$

$$= F(x) + \sum_{n=1}^{\infty} \sum_{y \in S_{m+n}} M(x, y)F(y) = F(x) + \sum_{n=1}^{\infty} M_{m,m+n} \sum_{y \in S_{m+n}} F(y)$$

$\square$
We now turn our attention to constant rate distributions for the partial order graph $(S, \preceq)$ and its covering graph $(S, \uparrow)$.

**Theorem 17.6.** Suppose that $X$ has constant rate $\alpha \in (0, 1)$ for $(S, \preceq)$. Then $X$ has density function $f$ given by

$$f(x) = \frac{\alpha(1-\alpha)^n}{(1-\alpha + \alpha k_0)(1-\alpha + \alpha k_1)\cdots(1-\alpha + \alpha k_n)}, \quad x \in S_n, \ n \in \mathbb{N}$$

*Proof.* As above, let $p_n = \mathbb{P}(X \in S_n) = \mathbb{P}(N = n)$ for $n \in \mathbb{N}$ so that $n \mapsto p_n$ is the density function of $N$. Let $P_n = \sum_{m=n+1}^{\infty} p_m$ for $n \in \mathbb{N}$, so that $n \mapsto P_n$ is the right probability function of $N$ for the strict partial order graph $(N, \prec)$. As noted earlier, the right probability function $F$ of $X$ for $(S, \preceq)$ is given by $F(x) = f(x) + P_n$ for $x \in S_n$ and $n \in \mathbb{N}$. Hence if $X$ has constant rate $\alpha \in (0, 1)$, then $f = \alpha F$ and so

$$F(x) = \frac{1}{1-\alpha} P_n, \quad x \in S_n, \ n \in \mathbb{N}$$

But $P_n - P_{n+1} = p_{n+1}$ for $n \in \mathbb{N}$ and moreover,

$$p_m = \mathbb{P}(X \in S_m) = \sum_{x \in S_m} f(x) = \sum_{x \in S_m} \alpha F(x) = \sum_{x \in S_m} \frac{\alpha}{1-\alpha} P_m = k_m \frac{\alpha}{1-\alpha} P_m, \quad m \in \mathbb{N}$$

Substituting we have $P_n - P_{n+1} = [\alpha k_{n+1}/(1-\alpha)]P_{n+1}$ for $n \in \mathbb{N}$ or equivalently,

$$P_{n+1} = \frac{1-\alpha}{1-\alpha + \alpha k_{n+1}} P_n, \quad n \in \mathbb{N}$$

Solving gives

$$P_n = \frac{(1-\alpha)^n}{(1-\alpha + \alpha k_0)\cdots(1-\alpha + \alpha k_n)} P_0, \quad n \in \mathbb{N}$$

Finally, $P_0 = 1 - p_0 = 1 - [\alpha/(1-\alpha)]P_0$ and so

$$P_0 = \frac{1-\alpha}{1-\alpha + \alpha k_0}$$

\qed

**Corollary 17.4.** Suppose that $X$ has constant rate $\alpha \in (0, 1)$ for $(S, \preceq)$ as in Theorem 17.6. Let $N$ denote the index variable of $X$, so that $N = n$ if and only if $X \in S_n$ for $n \in \mathbb{N}$.

(a) The probability density function of $N$ is given by

$$p_n = \frac{\alpha(1-\alpha)^n k_n}{(1-\alpha + \alpha k_0)(1-\alpha + \alpha k_1)\cdots(1-\alpha + \alpha k_n)}, \quad n \in \mathbb{N} \quad (17.2)$$

(b) The upper probability function of $N$ for $(N, \prec)$ is given by

$$P_n = \frac{(1-\alpha)^{n+1}}{(1-\alpha + \alpha k_0)\cdots(1-\alpha + \alpha k_n)}, \quad n \in \mathbb{N} \quad (17.3)$$

(c) The rate function $r$ of $N$ for $(N, \prec)$ is given by

$$r_n = \frac{\alpha k_n}{1-\alpha}, \quad n \in \mathbb{N}$$

(d) Given $N = n \in \mathbb{N}$, random variable $X$ is uniformly distributed on $S_n$.

So if $k_n$ is increasing, or decreasing, or constant in $n \in \mathbb{N}$, then $N$ has increasing rate, decreasing rate, or constant rate, respectively, for $(N, \prec)$. In the latter case, of course, $N$ has a geometric distribution. In the decreasing case, $k_n$ must eventually be constant in $n \in \mathbb{N}$. Equation (17.2) defines an interesting class of distributions on $\mathbb{N}$. Here are some special cases:
Example 17.4. If \( k_n = k \) for all \( n \in \mathbb{N} \), then as noted above, \( N \) has the geometric distribution on \( \mathbb{N} \) with success parameter \( \alpha k/(1 - \alpha + \alpha k) \), and has constant rate for each of the graphs \( (\mathbb{N}, \leq) \), \( (\mathbb{N}, <) \), and \( (\mathbb{N}, \uparrow) \). In particular, if \( k = 1 \), \( N \) has the geometric distribution with success parameter \( \alpha \), which of course must be the case since \((S, \succeq)\) is isomorphic to \((\mathbb{N}, \leq)\).

Example 17.5. If \( k_n = n + 1 \) for \( n \in \mathbb{N} \) then \( N \) has probability density given by

\[
p_n = \frac{\alpha(1 - \alpha)^n(n + 1)}{(1)(1 + \alpha)(1 + 2\alpha)\cdots(1 + n\alpha)}, \quad n \in \mathbb{N}
\]

It’s easy to see that \( p_{n+1} \geq p_n \) if and only if \( n \leq \sqrt{(1 - \alpha)/\alpha} \) and hence the distribution is unimodal with mode at \( \lceil\sqrt{(1 - \alpha)/\alpha}\rceil \). The rate function of \( N \) for \((\mathbb{N}, <)\) is given by \( r_n = \alpha(n + 1)/(1 - \alpha) \) for \( n \in \mathbb{N} \) so \( N \) has increasing rate.

Example 17.6. If \( \alpha = \frac{1}{2} \) then \( N \) has probability density function given by

\[
p_n = \frac{k_n}{(1 + k_0)(1 + k_1)\cdots(1 + k_n)}, \quad n \in \mathbb{N}
\]

The rate function of \( N \) for \((\mathbb{N}, <)\) is given by \( r_n = k_n \) for \( n \in \mathbb{N} \).

Example 17.7. If both \( k_n = n + 1 \) for \( x \in \mathbb{N} \) and \( \alpha = \frac{1}{2} \) then \( N \) has probability density function given by

\[
p_n = \frac{n + 1}{(n + 2)!} = \frac{1}{(n + 1)!} - \frac{1}{(n + 2)!}, \quad n \in \mathbb{N}
\]

The rate function of \( N \) for \((\mathbb{N}, <)\) is given by \( r_n = n + 1 \) for \( n \in \mathbb{N} \), so \( N \) has increasing rate.

Exercise 17.6. Find the probability generating function, mean, and variance for the distribution in Example 17.7.

17.7 Positive semigroups

In the general lexicographic construction considered in this chapter, only one case corresponds to a positive semigroup, and that case corresponds to \( k_0 = 1 \) and \( k_n = k \in \mathbb{N}_+ \) for \( n \in \mathbb{N}_+ \). We give a construction that leads to an elegant notation. Let \( I \) be a set with \( k \) elements, which we think of as a set of letters. We start with the free semigroup \((I^*, \cdot)\) as studied in Chapter 13, but then we impose the word equations \( ij = j^2 \) for \( i, j \in I \). It then follows more generally that if \( x \in I^* \) is a word of length \( n \in \mathbb{N}_+ \) ending in the letter \( i \in I \), then \( x = i^n \). Words of this form are distinct: if \( m, n \in \mathbb{N}_+ \) and \( i, j \in I \) then \( i^m = j^n \) if and only if \( i = j \) and \( m = n \). So our new semigroup consists of the set

\[
S = \{i^n : n \in \mathbb{N}, i \in I\}
\]

with the operation \( \cdot \) defined on \( S_+ = \{i^n : n \in \mathbb{N}_+, i \in I\} \) by

\[
i^m \cdot j^n = j^{m+n}, \quad i^m, j^n \in S_+
\]

Of course \( i^0 = e \) for \( i \in I \) where \( e \) is the empty word, which functions as the identity just as it does in \((I^*, \cdot)\).

The partial order \( \preceq \) associated with \((S, \cdot)\) is given by \( i^m \prec j^n \) if and only if \( m < n \). Hence \((S, \preceq)\) is a special case of the posets considered in the previous section, with \( S_0 = \{e\} \) and \( S_n = \{i^n : i \in I\} \) for \( n \in \mathbb{N}_+ \).

Theorem 17.7. Suppose that \( X \) is a random variable with values in \( S \) and that \( X \) is memoryless for \((S, \cdot)\). Then \( X \) has upper probability function \( F \) for \((S, \cdot)\) given by \( F(i^n) = q^n \) for \( i^n \in S \), where \( q \in (0, 1) \) is a parameter. The probability density function \( f \) of \( X \) is given as follows:

(a) If \( q < 1/(k - 1) \) then

\[
f(i^n) = \frac{(1 - q)q^n}{1 - q + kq}, \quad i^n \in S
\]
(b) If \(q \geq 1/(k-1)\) then
\[
f(i^n) = \frac{(1-q)q^n}{1-q+kq} + \left(\frac{1}{k-1}\right)^n \left[f(e) - \frac{1-q}{1-q+kq}\right], \quad i^n \in S
\]  
(17.5)
where \(0 \leq f(e) \leq 1-q\).

Proof. The memoryless property of an upper probability function \(F\) for \((S,\cdot)\) is
\[
F(i^m)F(j^n) = F(j^{m+n}), \quad i^m, j^n \in S_+
\]
So \(F(i^m) = F(j^{m+n})/F(j^n)\) and it follows that \(F(i^n)\) is constant in \(i \in I\) for each \(m \in \mathbb{N}_+\). Another application of the memoryless property then implies that \(F(i^n) = q^n\) for \(i^n \in S_+\) where \(q = F(i)\). Since \(F(e) = 1\), it follows that \(F(i^n) = q^n\) for all \(i^n \in S\).

Now let \(f\) denote the probability density function of \(X\). For \(n \in \mathbb{N}_+\) and \(i, j \in I\),
\[
F(i^n) = f(i^n) + \sum_{m=n+1}^{\infty} \sum_{u \in I} f(u^m)
\]
\[
F(j^n) = f(j^n) + \sum_{m=n+1}^{\infty} \sum_{u \in I} f(u^m)
\]
Since \(F(i^n)\) is constant in \(i \in I\) for each \(n \in \mathbb{N}_+\), it follows that \(f\) has this property as well. So, let \(\varphi(n)\) denote the common value of \(f(i^n)\) for \(i^n \in S_+\) and let \(\varphi(0) = f(e)\). It follows that
\[
q^n = \varphi(n) + k \sum_{m=n+1}^{\infty} \varphi(m), \quad n \in \mathbb{N}
\]
Subtracting the equation with \(n + 1\) from the equation with \(n\) gives
\[
(1-q)q^n = \varphi(n) + (k-1)\varphi(n+1)
\]
Using this result recursively gives (17.4). Finally, the only requirement on \(\varphi(0) = f(e)\) is that \(\varphi(n) \geq 0\) for \(n \in \mathbb{N}\). Solving the various inequalities gives \(f(e) = (1-q)/(1-q+kq)\) if \(q < 1/(k-1)\), and \(0 \leq f(e) \leq 1-q\) if \(q \geq 1/(k-1)\).

Corollary 17.5. Suppose again that \(X\) is a random variable with values in \(S\), Then \(X\) has an exponential distribution for \((S,\cdot)\) if and only if the probability density function \(f\) of \(X\) has the form
\[
f(i^n) = \frac{(1-q)q^n}{1-q+kq}, \quad i^n \in S
\]
where \(q \in (0,1)\) is a parameter. The upper probability function \(F\) of \(X\) for \((S,\cdot)\) is given by \(F(i^n) = q^n\) for \(i^n \in S_+\) and \(X\) has constant rate \((1-q)/(1-q+kq)\) for \((S,\cdot)\).

Proof. Random variable \(X\) has an exponential distribution for \((S,\cdot)\) if an only if \(X\) is memoryless and has constant rate with respect to counting measure \# (the only left-invariant measure, up to multiplication by positive constants). The result then follows from Theorems 17.6 and 17.7.

In the case \(q < 1/(k-1)\), the only memoryless distribution is the exponential distribution. In the case \(q \geq 1/(k-1)\), the exponential probability density function corresponds to choosing \(f(e) = (1-q)/(1-q+kq)\) in (17.5). But of course, this choice of \(f(e)\) is not the only possible one. Indeed, any choice of \(f(e)\) with \(0 \leq f(e) \leq 1-q\) will lead to a probability density function whose corresponding upper probability function \(F\) is \(F(i^n) = q^n\) for \(i^n \in S\), corresponding to a memoryless distribution. In particular, the upper probability function does not uniquely specify the distribution, and there are probability distributions that are memoryless but not exponential. The following exercise explores a concrete example.

Exercise 17.7. Suppose that \(k = 4\) and that \(q = \frac{1}{2}\), so that all of the distributions defined by (17.5) are memoryless for \((S,\cdot)\), with upper probability function \(F\) given by \(F(i^n) = \left(\frac{1}{2}\right)^n\) for \(i^n \in S\). Find the probability density function \(f\) for each of the following choices of \(f(e)\): \(f(e) = 0\), \(f(e) = \frac{1}{2}\), \(f(e) = \frac{2}{5}\), \(f(e) = \frac{1}{2}\).
On the other hand, for this positive semigroup, the constant rate property implies the full exponential property.

**Corollary 17.6.** Suppose again that $X$ is a random variable with values in $S$. If $X$ has constant rate for $(S, \preceq)$ then $X$ is exponential for $(S, \cdot)$.

**Proof.** Suppose that $F$ is the upper probability function of a distribution which has constant rate $\alpha \in (0, 1)$ for $(T, \preceq)$. Then from (??),

$$F(i^n) = \frac{(1 - \alpha)^n}{(1)(1 - \alpha + \alpha k)^n} = \left( \frac{1 - \alpha}{1 - \alpha + \alpha k} \right)^n, \quad i^n \in S$$

Hence the distribution is memoryless as well, and hence exponential for $(T, \cdot)$. □

So to review, every distribution with constant rate for $(S, \preceq)$ is memoryless (and hence exponential) for $(S, \cdot)$, but conversely, there are memoryless distributions that do not have constant rate. Of course, if $k = 1$ then $(S, \cdot)$ is isomorphic to $(\mathbb{N}, +)$ and the distribution defined by $g$ in (17.4) is the geometric distribution with rate $\alpha$.

### 17.8 Quotient Spaces

We continue with the semigroup $(S, \cdot)$ in the last section to illustrate some of the results in Chapter 9 on quotient spaces. Fix $j \in I$ and let $S_j = \{j^n : n \in \mathbb{N}\}$ so that $(S_j, \cdot)$ is the complete sub-semigroup of $(S, \cdot)$ generated by the element $j$. The corresponding quotient space is

$$S/S_j = \{e\} \cup (I - \{j\})$$

The basic assumptions are satisfied, so that $i^n \in S$ has a unique factoring over $S_j$ and $S/S_j$. Specifically, the non-trivial factoring is

$$i^n = j^{n-1}i, \quad n \in \mathbb{N}, \quad i \in I - \{j\}$$

If $X$ is a random variable with values in $S$ then $X = j^NY$ where $N$ takes values in $\mathbb{N}$ and $Y$ takes values in $S/S_j$. Let $f$ denote the probability density function of $X$. Then the probability density function of $(N,Y)$ is given by

$$P(N = n, Y = e) = f(j^n), \quad n \in \mathbb{N}$$

$$P(N = n, Y = i) = f(i^{n+1}), \quad n \in \mathbb{N}, \quad i \in I - \{j\}$$

**Theorem 17.8.** Suppose that $X$ has the exponential distribution given in Corollary 17.5. Then

(a) $N$ has the geometric distribution on $\mathbb{N}$ with success parameter $p = 1 - q$.

(b) $Y$ has probability density function $g$ given by

$$g(e) = \frac{1}{1 - q + kq}$$

$$g(i) = \frac{q}{1 - q + kq}, \quad i \in I - \{j\}$$

(c) $N$ and $Y$ are independent.

**Proof.** The fact that $N$ has an exponential distribution on $(\mathbb{N}, +)$ and that $N$ and $Y$ are independent follows from the general theory, specifically Theorem 9.3. The particular details in this theorem follow from the density function of $(N,Y)$ and standard techniques.

$$P(N = n, Y = e) = P(X = j^n) = \frac{(1 - q)q^n}{1 - q + kq}, \quad n \in \mathbb{N}$$

$$P(N = n, Y = i) = P(X = i^{n+1}) = \frac{(1 - q)q^{n+1}}{1 - q + kq}, \quad n \in \mathbb{N}, \quad i \in I - \{j\}$$

□
Our next discussion gives a counterexample about conditional exponential distributions. We allow the set \( I \) to be countably infinite.

**Theorem 17.9.** Let \( p_i \in (0, 1) \) for each \( i \in I \) and assume that \( \sum_{i \in I}(1-p_i)/p_i < \infty \). Suppose that random variable \( X \in S \) has probability density function \( f \) given by \( f(i^n) = d(1-p_i)^n \) for \( i^n \in S \) where

\[
d = \frac{1}{1 + \sum_{i \in I}(1-p_i)/p_i}
\]

Then the conditional distribution of \( X \) given \( X \in S \) is exponential for \((S_x, \cdot)\) for each \( x \in S \), but \( X \) does not have a memorryless distribution for \((S, \cdot)\) unless \( I \) is finite and \( p_i \) is constant in \( i \in I \).

**Proof.** First we verify that \( f \) is a valid density function on \( S \):

\[
\sum_{x \in S} f(x) = f(e) + \sum_{i \in I} \sum_{n=1}^{\infty} f(i^n) = d + d \sum_{i \in I} \sum_{n=1}^{\infty} (1-p_i)^n = d \left[ 1 + \sum_{i \in I}(1-p_i)/p_i \right] = 1
\]

by definition of \( d \). Next note that need only prove the theorem when \( x = i \in I \), since every other semigroup \( S_x \) is a sub-semigroup of one of these. Towards that end,

\[
\mathbb{P}(X \in S_i) = \sum_{n=0}^{\infty} f(i^n) = d + d \sum_{n=1}^{\infty} (1-p_i)^n = d \left[ 1 + (1-p_i)/p_i \right] = d/p_i
\]

Hence

\[
\mathbb{P}(X = i^n \mid X \in S_i) = \frac{f(i^n)}{\mathbb{P}(X \in S_i)} = p_i(1-p_i)^n, \quad n \in \mathbb{N}
\]

which is the geometric distribution with rate \( p_i \) under the isomorphism \( i^n \mapsto n \) for \( n \in \mathbb{N} \). But clearly \( X \) does not have a memoryless distribution for \((S, \cdot)\) unless \( S \) is finite and \( p_i = p \in (0, 1) \) for \( i \in I \). In this case, \( d = p/[p + \#(I)(1-p)] \) and so \( f(i^n) = d(1-p)^n \) for \( i^n \in S \), which agrees with the exponential distribution in Corollary 17.5.

**Theorem 17.10.** Suppose that \( X \) has the distribution in 17.9. Fix \( j \in I \) and consider the decomposition \( X = j^N Y \) where \( N \in \mathbb{N} \) and \( Y \in S/S_j \). Then \((N,Y)\) has probability density function defined as follows:

\[
\mathbb{P}(N = n, Y = e) = \frac{(1-p_j)^n}{1 + \sum_{y \in I}(1-p_y)/p_y}, \quad n \in \mathbb{N}
\]

\[
\mathbb{P}(N = n, Y = i) = \frac{(1-p_j)^{n+1}}{1 + \sum_{y \in I}(1-p_y)/p_y}, \quad n \in \mathbb{N}, \ i \in I - \{j\}
\]

**Proof.** First

\[
\mathbb{P}(N = n, Y = e) = \mathbb{P}(X = j^n) = d(1-p_j)^n, \quad n \in \mathbb{N}
\]

Next,

\[
\mathbb{P}(N = n, Y = i) = \mathbb{P}(X = i^{n+1}) = d(1-p_i)^{n+1}, \quad n \in \mathbb{N}, \ i \in I - \{j\}
\]

**Corollary 17.7.** Suppose again that \( X \) has the distribution in 17.9. Fix \( j \in I \) and consider the decomposition \( X = j^N Y \) where \( N \in \mathbb{N} \) and \( Y \in S/S_j \). Then

(a) \( N \) has probability density function given by

\[
\mathbb{P}(N = n) = \frac{(1-p_j)^n + \sum_{i \in I - \{j\}}(1-p_i)^{n+1}}{1 + \sum_{i \in I}(1-p_i)/p_i}, \quad n \in \mathbb{N}
\]

(b) \( Y \) has probability density function given by

\[
\mathbb{P}(Y = e) = \frac{1 + (1-p_j)/p_j}{1 + \sum_{i \in I}(1-p_i)/p_i}
\]

\[
\mathbb{P}(Y = i) = \frac{(1-p_i)/p_i}{1 + \sum_{y \in I}(1-p_y)/p_y}, \quad i \in I - \{j\}
\]

In particular, note from the last two results that \( N \) and \( Y \) are dependent.
Chapter 18

Miscellaneous Examples

18.1 Remaining Life Chain

Define the discrete graph \((N, \rightarrow)\) where the relation \(\rightarrow\) on \(N\) is defined by \(x \rightarrow (x - 1)\) for \(x \in N_+\) and \(0 \rightarrow x\) for all \(x \in N\). Thus, \(\rightarrow\) is the “leads-to” relation in the remaining life Markov chain, which in turn is the time-reversal of the success-runs chain that we will consider in the next section. For \(u : N \rightarrow \mathbb{R}\), the left operator is given by

\[
(uK)(x) = u(0) + u(x + 1)
\]

The left walk function \(\gamma_n\) of order \(n \in N\) is given by \(\gamma_n(x) = 2^n\) for \(x \in N\).

Suppose now that \(X\) is a random variable with values in \(N\) and with density function \(f\). The right probability function \(F\) of \(X\) for \((N, \rightarrow)\) is given by \(F(x) = f(x - 1)\) for \(x \in N_+\) and \(F(0) = \mathbb{P}(X \in N) = 1\). We have our usual support assumption that \(F(x) > 0\) for \(x \in N\), and this in turn means that we require \(f(x) > 0\) for \(x \in N\). The right rate function \(r\) of \(X\) for \((N, \rightarrow)\) is given by \(r(x) = f(x)/f(x - 1)\) for \(x \in N_+\) and \(r(0) = f(0)\). Note that both \(F\) and \(r\) determines the distribution of \(X\).

**Theorem 18.1.** The only constant rate distribution for \((N, \rightarrow)\) has rate \(1/2\) and is the geometric distribution on \(N\).

**Proof.** To have right constant rate \(\alpha \in (0, 1)\), \(f\) must satisfy \(f(0) = \alpha\) and \(f(x) = \alpha f(x - 1)\) for \(x \in N_+\). Thus \(f(x) = \alpha^{x+1}\) for \(x \in N_+\). But

\[
\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \alpha^{x+1} = \frac{\alpha}{1 - \alpha} = 1
\]

so \(\alpha = \frac{1}{2}\). Thus \(f(x) = \left(\frac{1}{2}\right)^{x+1}\) and \(F(x) = \left(\frac{1}{2}\right)^x\) for \(x \in N\).

Suppose again that \(X\) has density function \(f\). The random walk \(Y = (Y_1, Y_2, \ldots)\) on \((N, \rightarrow)\) associated with the distribution of \(X\) has transition probability function \(P\) given by

\[
P(x, x - 1) = \frac{f(x - 1)}{F(x)} = 1, \quad x \in N_+
\]

\[
P(0, x) = = \frac{f(x)}{F(0)} = f(x), \quad x \in N
\]

Thus, \(Y\) is a remaining life chain, with lifetime distribution \(f\). The chain is recurrent. The function \(h\) given by

\[
h(x) = \mathbb{P}(X \geq x) = \sum_{y=x}^{\infty} f(y), \quad x \in N
\]

is invariant for \(Y\). Note that \(h\) is the right probability function of \(X\) for \((N, \leq)\). So the chain is positive recurrent if and only if \(m := \mathbb{E}(X) < \infty\), in which case the normalizing constant is

\[
\sum_{x=0}^{\infty} \mathbb{P}(X \geq x) = 1 + m
\]
and the invariant density function $g$ is given by $g(x) = h(x)/(1 + m)$, $x \in \mathbb{N}$. Note that $g$ and $1 + m$ are the density function and mean, respectively, of $1 + X$, which is the return time to 0, starting at 0. In particular, the chain corresponding to the constant rate distribution is positive recurrent, and in this case, $m = 1$ and $g = f$. Finally, note that every remaining life chain is a random walk on $(\mathbb{N}, \rightarrow)$ corresponding to some distribution on $\mathbb{N}$.

Fix $n \in \mathbb{N}_+$. Note that a walk $y = (y_1, y_2, \ldots, y_n)$ of length $n$ in $(\mathbb{N}, \rightarrow)$ must start with a subsequence decreasing by 1 from term to term, perhaps to 0, and if so, followed by another subsequence of the same form, and so forth. Now let $k \in \mathbb{N}$ denote the number of times 0 occurs in $y$. Then let $j_1 = 1$ and $j_i$ the index following the $i$th 0 for $i \in \{1, 2, \ldots, k\}$. Then the probability density function of $(Y_1, Y_2, \ldots, Y_n)$ at $y$ has the form

$$f(y_{j_1}, f(y_{j_2}), \ldots, f(y_{j_k}))$$

This makes sense. The only source of randomness in the sequence $Y$ are the “rebirth” terms following the 0s. For the density function $f_n$ of $Y_n$, we use the general formula $f_n(y) = v_{n-1}(y)f(y)$ for $y \in S$. This leads to the following, each for $y \in \mathbb{N}$:

$$f_1(y) = f(y)$$
$$f_2(y) = f(0)f(y) + f(y + 1)$$
$$f_3(y) = f^2(0)f(y) + f(0)f(y + 1) + f(1)f(y) + f(y + 2)$$

The next term is

$$f_4(y) = f^3(0)f(y) + f^2(0)f(y + 1) + f(0)f(y + 2) + f(0)f(1)f(y) + f(1)f(0)f(y) + f(1)f(y + 1) + f(2)f(y) + f(y + 3)$$

In general, $f_n(y)$ has $2^{n-1}$ terms, one for each of the $2^{n-1}$ walks of length $n - 1$ in $(\mathbb{N}, \rightarrow)$ terminating in $y$. Recursively, for $n \in \mathbb{N}$,

$$f_{n+1}(y) = \sum_{x=0}^{n-1} f(x)f_{n-x}(y) + f(y + n), \quad y \in \mathbb{N}$$

### 18.2 The Success Runs Chain

Consider the discrete graph $(\mathbb{N}, \rightarrow)$ where the relation $\rightarrow$ on $\mathbb{N}$ is defined by $x \rightarrow 0$ and $x \rightarrow x + 1$ for $x \in \mathbb{N}$. Thus $\rightarrow$ is the “leads to” relation in the success run chain, the time reversal of the remaining life chain considered in Section 18.1, and so the reversal of the relation for the remaining life graph. The left operator $K$ is given by

$$(uK)(x) = u(x - 1), \quad x \in \mathbb{N}_+; \quad (uK)(0) = \sum_{y=0}^{\infty} u(y)$$

for $u : S \rightarrow \mathbb{R}$, assuming that the series for $(uK)(0)$ converges. In particular, the left walk function $\gamma$ for $(\mathbb{N}, \rightarrow)$ is given by $\gamma(x) = 1$ for $x \in \mathbb{N}_+$ while $\gamma(0) = \infty$. More generally, for $n \in \mathbb{N}_+$, $\gamma_n(x) = 1$ for $x \in \mathbb{N}_+$ and $\gamma_n(0) = \infty$.

Suppose now that $X$ is a random variable with values in $\mathbb{N}$ and with density function $f$. The right probability function $F$ of $X$ for $(\mathbb{N}, \rightarrow)$ is given by

$$F(x) = f(0) + f(x + 1), \quad x \in \mathbb{N}$$

Our support assumption that $F(x) > 0$ for $x \in \mathbb{N}$ implies that either $f(0) > 0$ or $f(x) > 0$ for all $x \in \mathbb{N}_+$. The right rate function $r$ of $X$ for $(\mathbb{N}, \rightarrow)$ is given by

$$r(x) = \frac{f(x)}{f(0) + f(x + 1)}, \quad x \in \mathbb{N}$$

Note that $F$ completely determines the distribution. Given a valid right probability function $F$, we can recover $f(0) = \lim_{x \rightarrow \infty} F(x)$, and then $f(x + 1) = F(x) - f(0)$ for $x \in \mathbb{N}$. On the other hand, what
conditions on \( F \) guarantee that \( F \) is a right probability function for \((N, \rightarrow)\)? It would seem that there must exist \( a \in (0, 1) \) with \( F(x) > a \) for all \( x \in \mathbb{N} \), and

\[
\sum_{x=0}^{\infty} [F(x) - a] = 1 - a
\]

Then \( f \) given by \( f(0) = a \) and \( f(x+1) = F(x) - a \) for \( x \in \mathbb{N} \) is a density function with right probability function \( F \). There are no distributions with right constant rate since \( \gamma(0) = \infty \).

**Example 18.1.** Suppose that \( f \) is the geometric probability density function on \( \mathbb{N} \) with success parameter \( p \), so that \( f(x) = p(1-p)^x \) for \( x \in \mathbb{N} \). The rate function \( r \) of \( f \) for \((\mathbb{N}, \rightarrow)\) reduces to

\[
r(x) = \frac{(1-p)^x}{1+(1-p)^{x+1}}, \quad x \in \mathbb{N}
\]

So \( r(x) \) decreases to 0 as \( x \) increases to \( \infty \).

Suppose again that \( X \) is a random variable with values in \( \mathbb{N} \), density function \( f \), and right probability function \( F \) for \((\mathbb{N}, \rightarrow)\). Let \( Y = (Y_1, Y_2, \ldots) \) denote the random walk on \((\mathbb{N}, \rightarrow)\) associated with the distribution of \( X \). So the transition probability function \( P \) given by

\[
P(x, x+1) = \frac{f(x+1)}{f(0) + f(x+1)}, \quad P(x, 0) = \frac{f(0)}{f(0) + f(x+1)}; \quad x \in \mathbb{N}
\]

so that \( Y \) is a success-runs chain. At first let’s consider the random walk on \((\mathbb{N}, \rightarrow)\) with transition probability function \( P \), but without a specified initial distribution. For the main theorem that follows, let \( T_0 \) denote the return time to 0, starting at 0, and let \( G \) be the right probability function of \( T_0 \) for the graph \((\mathbb{N}, <)\):

\[
G(y) = P(T_0 > y) = \prod_{x=0}^{y-1} P(x, x+1) = \prod_{x=0}^{y-1} \frac{f(x+1)}{f(0) + f(x+1)} = \prod_{x=1}^{y} \frac{f(x)}{f(0) + f(x)}, \quad y \in \mathbb{N}
\]

**Theorem 18.2.** The success-runs chain with transition probability probability function \( P \) is irreducible, aperiodic, and positive recurrent, with invariant density function \( g \) given by

\[
g(y) = \frac{G(y)}{\sum_{x=0}^{\infty} G(x)}, \quad y \in \mathbb{N}
\]

*Proof.* In general, success-runs chains are irreducible and aperiodic. The function \( t \mapsto t/[f(0) + t] \) on \([0, 1]\) is increasing (and concave downward), so

\[
\frac{f(x)}{f(0) + f(x)} < \frac{1}{f(0) + 1} < 1, \quad x \in \mathbb{N}_+
\]

Hence

\[
G(y) \leq \left( \frac{1}{f(0) + 1} \right)^y, \quad y \in \mathbb{N}
\]

so \( E(T_0) = \sum_{y=0}^{\infty} G(y) < \infty \) and the chain is positive recurrent. From the general theory of success-runs chains, \( G \) is invariant for the chain, and so the normalized function \( g \) is the invariant density function. \( \square \)

From the general theory of reliability chains, the time reversal of the success-runs chain here is the remaining life chain studied in Section 18.1, corresponding to the density function \( \hat{f} \) on \( \mathbb{N} \) given by

\[
\hat{f}(x) = \prod_{y=0}^{x-1} \frac{f(y+1)}{f(0) + f(y+1)} - \prod_{y=0}^{x} \frac{f(y+1)}{f(0) + f(y+1)} = \prod_{y=1}^{x} \frac{f(y)}{f(0) + f(y)} - \prod_{y=1}^{x+1} \frac{f(y)}{f(0) + f(y)}, \quad x \in \mathbb{N}
\]

Needless to say, \( \hat{f} \neq f \) in general, so the time reversal of the success-runs chain corresponding to density function \( f \) is not the remaining life chain corresponding to density \( f \).

**Problem 18.1.** Is it ever the case that \( \hat{f} = f \)?
Clearly also, not every success-runs chain is a random walk on \((\mathbb{N}, \rightarrow)\) corresponding to a distribution on \(\mathbb{N}\). This may be the case even when the success-runs chain is positive recurrent.

**Example 18.2.** Consider the success-runs chain with transition probability function \(P\) given by

\[
P(x, x + 1) = a(x), \quad P(x, 0) = 1 - a(x), \quad x \in \mathbb{N}
\]

where \(a : \mathbb{N} \to (0, 1)\). Our goal is to find a density function \(f\) on \(\mathbb{N}\) with

\[
a(x) = \frac{f(x + 1)}{f(0) + f(x + 1)}, \quad x \in \mathbb{N}
\]

Equivalently,

\[
f(x + 1) = \frac{a(x)}{1 - a(x)} f(0), \quad x \in \mathbb{N}
\]

Thus, the density function \(f\) exists if and only if

\[
c := \sum_{x=0}^{\infty} \frac{a(x)}{1 - a(x)} < \infty
\]

in which case

\[
f(0) = \frac{1}{1 + c}, \quad f(x + 1) = \frac{a(x)}{(1 + c)(1 - a(x))}, \quad x \in \mathbb{N}
\]

### 18.3 The Wheel

Consider the (undirected) wheel graph \((S, \rightarrow)\) on \(m + 1\) vertices where \(S = \{0, 1, \ldots, m\}\) for \(m \in \{3, 4, \ldots\}\). Vertex 0 is the center and the vertices \(\{1, 2, \ldots, m\}\) form a cycle on the rim of the wheel. That is, \(0 \rightarrow x\) and \(x \rightarrow 0\) for \(x \in \{1, 2, \ldots, m\}\). In addition, \(x \rightarrow x + 1\) and \(x \rightarrow x - 1\) for \(x \in \{1, 2, \ldots, m\}\) where we interpret the addition and subtraction modulo \(m\) (so \(m + 1 = 1\) and \(1 - 1 = m\)).

**Theorem 18.3.** The left walk function \(\gamma_n\) of order \(n \in \mathbb{N}\) for \((S, \rightarrow)\) is given by

\[
\gamma_n(0) = \left(\frac{1}{2} + \frac{1}{\sqrt{m + 1}}\right) m \left(1 + \sqrt{m + 1}\right)^{n-1} + \left(\frac{1}{2} - \frac{1}{\sqrt{m + 1}}\right) m \left(1 - \sqrt{m + 1}\right)^{n-1}
\]

\[
\gamma_n(x) = \left(\frac{1}{2} + \frac{1}{\sqrt{m + 1}}\right) \left(1 + \sqrt{m + 1}\right)^n + \left(\frac{1}{2} - \frac{1}{\sqrt{m + 1}}\right) \left(1 - \sqrt{m + 1}\right)^n, \quad x \in \{1, 2, \ldots, m\}
\]

**Proof.** Let \(a_n = \gamma_n(0)\). By symmetry, \(\gamma_n\) is constant on \(\{1, 2, \ldots, m\}\) so let \(b_n\) denote the common value. Starting at 0, the next vertex on a walk must be on the rim. Starting at a vertex on the rim, the next vertex on a walk must be one of the neighbors on the rim or the center. Thus \(a_{n+1} = mb_n\) and \(b_{n+1} = 2b_n + a_n\) for \(n \in \mathbb{N}\), subject to the initial conditions \(a_0 = b_0 = 1\). Substituting we get the second order difference equation

\[
b_{n+1} = 2b_n + mb_{n-1}, \quad n \in \mathbb{N}_+
\]

subject to the initial conditions \(b_0 = 1, b_1 = 3\). The characteristic equation is \(r^2 - 2r - m = 0\) with solutions \(r = 1 \pm \sqrt{m + 1}\). So the general solution is

\[
b_n = c_1 \left(1 + \sqrt{m + 1}\right)^n + c_2 \left(1 - \sqrt{m + 1}\right)^n, \quad n \in \mathbb{N}
\]

Applying the initial conditions gives the expression for \(b_n\) in the theorem. Then returning to \(a_n = mb_{n-1}\) for \(n \in \mathbb{N}_+\) gives the expression for \(a_n\). \(\Box\)

**Theorem 18.4.** Suppose that \(X\) is a random variable with values in \(S\) and with density function \(f\). The right probability function \(F\) of \(X\) for \((S, \rightarrow)\) is given by

\[
F(0) = \sum_{x=1}^{m} f(x) = 1 - f(0)
\]

\[
F(x) = f(x - 1) + f(x + 1) + f(0), \quad x \in \{1, 2, \ldots, m\}
\]

The right probability function \(F\) determines the distribution except when \(m\) is a multiple of 4.
Proof. The formula for $F$ follows directly from the definition, and of course $F = Kf$ where $K$ is the adjacency matrix. It’s known that the spectrum of $K$ consists of $1 \pm \sqrt{m+1}$ together with $2\cos(2k\pi/m)$ for $k \in \{1, 2, \ldots, m-1\}$. So 0 is an eigenvalue of $L$ if and only if $2k\pi/m = \pi/2$ for some $k \in \{1, 2, \ldots, m-1\}$, or equivalently $k = m/4$, so that $m$ is a multiple of 4. In this case 0 is an eigenvalue of multiplicity 2, corresponding to $k = m/4$ and $k = 3m/4$. An eigenfunction $g$ corresponding to 0 is given by $g(2x) = 0$ for $x \in \{0, 1, \ldots, m/2\}$ and $g(2x+1) = (-1)^x$ for $x \in \{0, 1, \ldots, m/2-1\}$. So the result follows from Corollary 4.1. 

Theorem 18.5. The unique constant rate distribution for $(S, \rightarrow)$ has rate 

$$\alpha = \frac{\sqrt{m+1} - 1}{m}$$

and probability density function $f$ given by 

$$f(0) = \frac{\sqrt{m+1} - 1}{m-1 + \sqrt{m+1}},$$

$$f(x) = \frac{1}{m-1 + \sqrt{m+1}}, \quad x \in \{1, 2, \ldots, m\}$$

Proof. Suppose that the density function $f$ has constant rate $\alpha \in (0, \infty)$. Let $f(0) = p \in (0, 1)$. By symmetry, $f$ will be constant on the rim so let $f(x) = q$ for $x \in \{1, 2, \ldots, m\}$. We have $p + mq = 1$ so $q = (1 - p)/m$. The constant rate conditions are

$$p = \alpha(1 - p)$$

$$q = \alpha(p + 2q)$$

Solving gives $p = \alpha/(1 + \alpha)$ and

$$q = \frac{\alpha}{1 - 2\alpha}p = \frac{\alpha^2}{(1 + \alpha)(1 - 2\alpha)}$$

The condition $p + mq = 1$ then leads to the equation

$$m\alpha^2 + 2\alpha - 1 = 0$$

with the unique solution in $(0, 1)$ given by

$$\alpha = \frac{\sqrt{m+1} - 1}{m}$$

Then the constant rate density function $f$ is given by

$$f(0) = \frac{\sqrt{m+1} - 1}{m-1 + \sqrt{m+1}},$$

$$f(x) = \frac{(\sqrt{m+1} - 1)^2}{(m-1 + \sqrt{m+1})(m + 2 - 2\sqrt{m+1})}, \quad x \in \{1, 2, \ldots, m\}$$

Proof. For another proof, it’s well known that the largest eigenvalue of $K$ is $1 + \sqrt{m+1}$ with eigenfunction $g$ given by $g(0) = \sqrt{m+1} - 1$ and $g(x) = 1$ for $x \in \{1, 2, \ldots, m\}$. Hence the constant rate is

$$\alpha = \frac{1}{1 + \sqrt{m+1}} = \frac{\sqrt{m+1} - 1}{m}$$

The corresponding density function is obtained by normalizing $g$. Since $c = \sum_{x \in S} g(x) = m - 1 + \sqrt{m+1}$, the function $f(x) = g(x)/c$ for $x \in S$ gives the density function in the theorem. 

Example 18.3. In the special case that $m = 3$ we get $\alpha = 1/3$ and $f(x) = 1/4$ for $x \in S$. But we already knew this since with $m = 3$, the graph $(S, \rightarrow)$ is regular with degree 3.
Example 18.4. In the special case $m = 4$ we get $\alpha = (\sqrt{5} - 1)/4 \approx 0.309$. For the constant rate density function, $f(0) = \sqrt{5} - 2$ and $f(x) = (3 - \sqrt{5})/4$ for $x \in \{1, 2, 3, 4\}$.

Theorem 18.6. Suppose that $X_m$ has the constant rate distribution on the wheel graph of order $m$. Then the distribution of $X_m/m$ converges to the uniform distribution on $[0, 1]$ as $m \to \infty$.

Proof. Let $p_m$ denote the common value of $f_m(k)$ for $k \in \{1, 2, \ldots, m\}$ as given above. So $f(0) = 1 - mp_m$. Then for $x \in [0, 1]$,

$$P(X_m/m \leq x) = \sum_{k=0}^{m} 1(k/m \leq x) f_m(k) = f_m(0) + p_m \sum_{k=1}^{m} 1(k/m \leq x)$$

$$= f_m(0) + mp_m \frac{1}{m} \sum_{k=1}^{m} 1(k/m \leq x)$$

But as $m \to \infty$, $f_m(0) \to 0$, $mp_m \to 1$, and $\frac{1}{m} \sum_{k=1}^{m} 1(k/m \leq x) \to \int_{0}^{x} 1dt = x$. $\Box$

Theorem 18.7. The random walk on $(S, \rightarrow)$ associated with the constant rate distribution has transition matrix $P$ given as follows: For $x \in \{1, 2, \ldots, m\}$,

$$P(x, 0) = \frac{(\sqrt{m+1} - 1)^2}{m}$$

$$P(0, x) = \frac{1}{m}$$

$$P(x, y) = \frac{\sqrt{m+1} - 1}{m}, \quad y \in \{x-1, x+1\}$$

where $x - 1$ and $x + 1$ (mod $m$) are the neighbors of $x$ on the cycle.

Proof. This follows from the results above and the definition of $P$ as $P(x, y) = f(y)/F(x)$ for $x \to y$. $\Box$
Part IV

Appendices
Appendix A

Characterizations of the Poisson Distribution

Consider the partial order graph \((S, \subseteq)\) of Chapter 16, where \(S\) is the collection of finite subsets of \(\mathbb{N}\). Suppose that \(X\) is a random variable with values in \(S\) and that \(X\) has constant rate \(\alpha \in (0, 1)\) for \((S, \subseteq)\). We do not know if \(X\) exists, but if it does, we showed that \(N = \#(X)\) satisfies

\[
\alpha \mathbb{P}(N = k) = \mathbb{E} \left[ (-1)^{N-k} \binom{N}{k} \right], \quad k \in \mathbb{N} \tag{A.1}
\]

Equation (A.1) has several simple variations. The goal of this appendix is to show that all characterize the Poisson distribution.

**Theorem A.1.** The binomial function \(B\) defined by

\[
B(k, n) = \binom{n}{k}, \quad k, n \in \mathbb{N}
\]

is a kernel for the total order graph \((\mathbb{N}, \leq)\). Its inverse is given by

\[
B^{-1}(k, n) = (-1)^{n-k} \binom{n}{k}, \quad k, n \in \mathbb{N}
\]

**Proof.** The function \(B\) is a kernel for \((\mathbb{N}, \leq)\) since \(B(k, n) = 0\) for \(k > n\). It has an inverse since \(B(k, k) = 1 \neq 0\) for \(k \in \mathbb{N}\). The functions \(B\) and \(B^{-1}\) as given above are inverses with respect to the usual product operation:

\[
\sum_{j=k}^{n} (-1)^{n-j} \binom{n}{j} \binom{j}{k} = \mathbf{1}(k = n), \quad (k, n) \in \mathbb{N}^2
\]

As always, the kernel \(I\) given by \(I(k, n) = \mathbf{1}(k = n)\) for \((k, n) \in \mathbb{N}^2\) is the identity. \(\Box\)

Suppose that \(N\) is a random variable with values in \(\mathbb{N}\) and with probability density function \(f\). The condition (A.1) that we started with is

\[
\alpha f(k) = \mathbb{E} \left[ (-1)^{N-k} \binom{N}{k} \right] = \sum_{n=k}^{\infty} (-1)^{n-k} \binom{n}{k} f(n), \quad k \in \mathbb{N} \tag{A.2}
\]

with the underlying assumption that the series on the right converges absolutely. Equivalently, \(f\) is an eigenfunction of \(B^{-1}\) corresponding to the eigenvalue \(\alpha \in (0, \infty)\), on the space \(\mathcal{L}_1\). By the Möbius inversion formula in Corollary 2.2, equation (A.2) is equivalent to

\[
f(k) = \alpha \mathbb{E} \left[ \binom{N}{k} \right] = \alpha \sum_{n=k}^{\infty} \binom{n}{k} f(n), \quad k \in \mathbb{N} \tag{A.3}
\]

Equivalently, \(f\) is an eigenfunction of \(B\) corresponding to the eigenvalue \(1/\alpha\). We will soon collect other conditions that are equivalent, so we will refer to all of them simply as the equivalent conditions with parameter \(\alpha\). Note that \(\alpha = f(0)\).
Theorem A.2. Suppose that \( N \) has the Poisson distribution on \( \mathbb{N} \) with parameter \( \beta \in (0, \infty) \). Then \( N \) satisfies the equivalent conditions with parameter \( \alpha = e^{-\beta} \).

Proof. The computation is a familiar one, in a variety of contexts. It suffices to show that the function \( g \) is given by \( g(n) = \frac{\beta^n}{n!} \) for \( n \in \mathbb{N} \) is an eigenfunction of \( B \) corresponding to the eigenvalue \( e^\beta \).

\[
\sum_{n=k}^{\infty} \binom{n}{k} g(n) = \sum_{n=k}^{\infty} \frac{n\beta^n}{n!} = \frac{\beta^k}{k!} \sum_{n=k}^{\infty} \frac{\beta^{n-k}}{(n-k)!} = e^\beta \frac{\beta^k}{k!} = e^\beta g(k), \quad k \in \mathbb{N}
\]

Our goal is to show that the Poisson distribution is the only distribution on \( \mathbb{N} \) satisfying the equivalent conditions.

Theorem A.3. Suppose that \( N \) is a random variable with values in \( \mathbb{N} \) and with probability generating function \( P \). Then \( N \) satisfies the equivalent conditions with parameter \( \alpha \in (0, 1) \) if and only if

\[
P(t) = \alpha P(t + 1), \quad t \in \mathbb{R}
\]  

(A.4)

Proof. Let \( f \) denote the probability density function of \( f \). Recall that the probability generating function \( P \) is defined by

\[
P(t) = \mathbb{E}(t^N) = \sum_{n=0}^{\infty} f(n)t^n
\]

for \( t \) in the interval of absolute convergence, which must be at least \((-1, 1]\), since \( P(1) = 1 \). Suppose that \( N \) satisfies the equivalent conditions. Then using (A.3), a sum interchange, and the binomial theorem we have

\[
P(t) = \sum_{k=0}^{\infty} t^k f(k) = \sum_{k=0}^{\infty} t^k \alpha \sum_{n=k}^{\infty} \binom{n}{k} f(n)
\]

\[
= \alpha \sum_{n=0}^{\infty} f(n) \sum_{k=0}^{n} \binom{n}{k} t^k = \alpha \sum_{n=0}^{\infty} f(n) (1 + t)^n = \alpha P(t + 1)
\]

In addition, \( t + 1 \) must be in the interval of convergence, so this interval must be \( \mathbb{R} \). Conversely, suppose that \( P(t) = P(t + 1) \) for \( t \in \mathbb{R} \). Then using the same tricks as before,

\[
\sum_{k=0}^{\infty} t^k f(k) = \sum_{k=0}^{\infty} t^k \alpha \sum_{n=k}^{\infty} \binom{n}{k} f(n), \quad t \in \mathbb{R}
\]

Equating coefficients gives

\[
f(k) = \alpha \sum_{n=k}^{\infty} \binom{n}{k} f(n), \quad k \in \mathbb{N}
\]

The function \( t \mapsto P(t + 1) \) is sometimes called the factorial moment generating function since the derivatives at \( 0 \) give \( \mathbb{E}[N^{(k)}] \) for \( k \in \mathbb{N} \). So among the equivalent conditions is that the probability generating function be proportional to the factorial moment generating function. The following results will be useful in the proof of our main Theorem A.6 below.

Theorem A.4. Suppose that \( f \) is a probability density function satisfying the equivalent conditions with parameter \( \alpha \in (0, 1) \), and with probability generating function \( P \). Then for every \( r \in \mathbb{N}_+ \), the function \( f_r \) defined by

\[
f_r(n) = \alpha^{r-1} r^n f(n), \quad n \in \mathbb{N}
\]

is a probability density function satisfying the equivalent conditions with parameter \( \alpha^r \). The probability generating function \( P_r \) of \( f_r \) is given by \( P_r(t) = \alpha^{r-1} P(rt) \) for \( t \in \mathbb{R} \).
Proof. The generating function $P_r$ of $f_r$ is given by

$$P_r(t) = \sum_{n=0}^{\infty} f_r(n)t^n = \alpha^{-1} \sum_{n=1}^{\infty} f(n)r^n t^n = \alpha^{-1}P(rt), \quad t \in R$$

Using the condition in Theorem A.1,

$$P_r(1) = \alpha^{-1}P(r) = \alpha^{-1} \frac{1}{\alpha^r}P(0) = \frac{\alpha^r}{\alpha^r} = 1$$

Hence $P_r$ is a probability generating function and so $f_r$ is a probability density function. Finally, note that $P_r(0) = f_r(0) = \alpha^{-1}P(0) = \alpha^r$. Thus,

$$P_r(t) = \alpha^{-1}P(rt) = \alpha^{-1}\alpha^r P(rt + r) = \alpha^r \alpha^{-1}P[r(t + 1)] = \alpha^r P_r(t + 1), \quad t \in R$$

Note that $f_1 = f$. If any of the distributions are Poisson, then they all are.

**Corollary A.1.** If $f_r$ is a Poisson density function for some $r \in \mathbb{N}_+$ then $f_r$ is a Poisson density function for all $r \in \mathbb{N}_+$, with parameter $\beta r$ where $\beta = -\ln \alpha$.

**Proof.** Suppose that $f_r$ is Poisson for some $r \in \mathbb{N}_+$. Note that $f_r(0) = \alpha^{-1}f(0) = \alpha^r$. Hence the Poisson parameter is $\beta r$ where $\beta = -\ln \alpha$. So

$$f_r(n) = e^{-\beta r} \frac{(r\beta)^n}{n!} = e^{-(r-1)\beta} \frac{r^n}{n!} f(n), \quad n \in \mathbb{N}$$

It follows that $f(n) = e^{-\beta} \frac{r^n}{n!}$ for $n \in \mathbb{N}$ so that $f$ is the Poisson density function with parameter $\beta$. Then clearly $f_r$ is the Poisson density with parameter $\beta r$ for every $r \in \mathbb{N}_+$.

**Theorem A.5.** Suppose again that $f$ is a probability density function satisfying the equivalent conditions, with parameter $\alpha \in (0,1)$, and that $r \in \mathbb{N}_+$. If $(X,N)$ has values in $\mathbb{N}_+^2$ with probability density function $g_r$ given by

$$g_r(k,n) = \alpha^r r^k \binom{n}{k} f(n), \quad (k,n) \in \mathbb{N}_+^2$$

(A.5)

(a) $X$ has probability density function $f_r$.

(b) $N$ has probability density function $f_{r+1}$

(c) For $n \in \mathbb{N}$, the conditional distribution of $X$ given $N=n$ is binomial with parameters $n$ and $r/(r+1)$.

**Proof.** The proofs are straightforward.

(a) Using condition (A.3),

$$\sum_{n=k}^{\infty} g_r(k,n) = \alpha^r \sum_{n=k}^{\infty} \binom{n}{k} f(n) = \alpha^r \frac{1}{\alpha} f(k) = \alpha^{r-1}f(k) = f_r(k), \quad k \in \mathbb{N}$$

It follows that $g_r$ is a valid probability density function on $\mathbb{N}_+^2$ and that $X$ has probability density function $f_r$.

(b) Next,

$$\sum_{k=0}^{n} g_r(k,n) = \alpha^r f(n) \sum_{k=0}^{n} \binom{n}{k} r^k = \alpha^r (r+1)^n f(n) = f_{r+1}(n), \quad n \in \mathbb{N}$$

so $N$ has probability density function $f_{r+1}$.

(c) For $n \in \mathbb{N}$, the conditional density of $X$ given $N=n$ is

$$\frac{g_r(k,n)}{f_{r+1}(n)} = \frac{\alpha^r r^k \binom{n}{k} f(n)}{\alpha^r (r+1)^n f(n)} = \binom{n}{k} \left(\frac{r}{r+1}\right)^k \left(\frac{1}{r+1}\right)^{n-k}, \quad k \in \{0,1,\ldots,n\}$$
Theorem A.6. A probability distribution on $\mathbb{N}$ satisfies the equivalent conditions if and only if it is Poisson.

Proof. We already showed in Theorem A.2 that the Poisson distribution satisfies the equivalent conditions, so we just need to show the converse. We will use the Rao-Rubin characterization of the Poisson distribution: Suppose that $(X, N)$ takes values in $\mathbb{N}^2$. If the conditional distribution of $X$ given $N$ is binomial with parameters $N$ and $p$ and if the distribution of $X$ is the same as the conditional distribution of $X$ given $X = N$, then $N$ has a Poisson distribution.

Towards that end, suppose that $f$ satisfies the equivalent conditions and let $P$ denote the corresponding probability generating function. Let $r \in \mathbb{N}_+$ and let $(X, N)$ have the probability density function $g_r$ given in Theorem A.5. Using (A.4)

$$P(X = N) = \sum_{n=0}^{\infty} \alpha^r r^n \binom{n}{n} f(n) = \alpha^r P(r) = \alpha^r \frac{1}{\alpha^r} P(0) = \alpha$$

Hence

$$P(X = k \mid X = N) = \frac{P(X = k, X = N)}{P(X = N)} = \frac{P(X = k, N = k)}{P(X = N)} = \frac{\alpha^r r^k \binom{k}{k} f(k)}{\alpha} = \alpha^{r-1} r^k f(k) = f_r(k) = P(X = k), \quad k \in \mathbb{N}$$

It follows that $N$ has the Poisson distribution. From Corollary A.1, $f$ is the Poisson density function with parameter $\beta = -\ln \alpha$.

Thus, (A.2), (A.3), and (A.4) characterize the Poisson distribution. These conditions are analytically more elegant than most characterizations of the Poisson distribution, but the probabilistic interpretation is less clear.
Appendix B

Solutions to Exercises

Chapter 2

Exercise 2.1  By assumption, → as a subset of $S^2$ is in $\mathcal{F}^2$.
(a) The function $\varphi : S^2 \to S^2$ defined by $\varphi(x,y) = (y,x)$ is measurable, since it has measurable coordinate functions. As a subset of $S^2$, ← is the inverse image of → under $\varphi$, and hence is in $\mathcal{F}^2$.
(b) As a subset of $S^2$, ↑ is the complement of → and hence is in $\mathcal{F}^2$.

Exercise 2.2  By assumption, → and ↑ as subsets of $S^2$ are in $\mathcal{F}^2$. Hence
(a) The union of → and ↑ is in $\mathcal{F}^2$.
(b) The intersection of → and ↑ is in $\mathcal{F}^2$.
(c) The set difference of → and ↑ is in $\mathcal{F}^2$.

Exercise 2.3  For the diamond graph $(S,\leftrightarrow)$:
(a) The adjacency matrix is
$$L = \begin{bmatrix}
0 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 \\
1 & 1 & 1 & 0
\end{bmatrix}$$
(b) The eigenvalues of $L$ are $\frac{1}{2}(1 + \sqrt{17}), \frac{1}{2}(1 - \sqrt{17}), -1, 0$, with corresponding eigenvectors
$$\left(\frac{\sqrt{17} - 1}{4}, 1, \frac{\sqrt{17} - 1}{4}, 1\right), \left(-\frac{\sqrt{17} - 1}{4}, 1, -\frac{\sqrt{17} - 1}{4}, 1\right), (0, -1, 0, 1), (-1, 0, 1, 0)$$
(c) The walk function of order $n \in \mathbb{N}$ is given by
$$w_n(1) = w_n(3) = \frac{\sqrt{17} - 3}{2\sqrt{17}} \left(\frac{1 - \sqrt{17}}{2}\right)^n + \frac{\sqrt{17} + 3}{2\sqrt{17}} \left(\frac{1 + \sqrt{17}}{2}\right)^n$$
$$w_n(2) = w_n(4) = \frac{\sqrt{17} - 5}{2\sqrt{17}} \left(\frac{1 - \sqrt{17}}{2}\right)^n + \frac{\sqrt{17} + 5}{2\sqrt{17}} \left(\frac{1 + \sqrt{17}}{2}\right)^n$$
(d) The generating function $W$ is given by
$$W(1,t) = W(3,t) = \frac{\sqrt{17} - 3}{2\sqrt{17}} \frac{1}{2 - t(1 - \sqrt{17})} + \frac{\sqrt{17} + 3}{2\sqrt{17}} \frac{1}{2 - t(1 + \sqrt{17})}, \quad |t| < \frac{2}{1 + \sqrt{17}}$$
$$W(2,t) = W(4,t) = \frac{\sqrt{17} - 5}{2\sqrt{17}} \frac{1}{2 - t(1 - \sqrt{17})} + \frac{\sqrt{17} + 5}{2\sqrt{17}} \frac{1}{2 - t(1 + \sqrt{17})}, \quad |t| < \frac{2}{1 + \sqrt{17}}$$
Exercise 2.4 For the bull graph \((S, \leftrightarrow)\):

(a) The adjacency matrix is

\[
L = \begin{bmatrix}
0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 \\
1 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0
\end{bmatrix}
\]

(b) The eigenvalues of \(L\) are \(\frac{1}{2}(1 + \sqrt{3})\), \(\frac{1}{2}(1 - \sqrt{3})\), \(\frac{1}{2}(-1 + \sqrt{5})\), \(\frac{1}{2}(-1 - \sqrt{5})\), 0 with corresponding eigenvectors

\[
\begin{align*}
(2, \frac{1 + \sqrt{3}}{2}, \frac{1 + \sqrt{3}}{2}, 1, 1), & \quad \left(0, \frac{1 + \sqrt{5}}{2}, -\frac{1 - \sqrt{5}}{2}, -1, -1\right), \\
(2, \frac{1 - \sqrt{3}}{2}, \frac{1 - \sqrt{3}}{2}, 1, 1), & \quad \left(0, \frac{1 - \sqrt{5}}{2}, -\frac{1 + \sqrt{5}}{2}, -1, -1\right), (-1, 0, 0, 1, 1)
\end{align*}
\]

(c) The walk function of order \(n \in \mathbb{N}\) is given by

\[
\begin{align*}
w_n(1) &= \frac{\sqrt{3} - 5}{\sqrt{3}} \left(\frac{1 - \sqrt{3}}{2}\right)^{n-1} + \frac{\sqrt{3} + 5}{\sqrt{3}} \left(\frac{1 + \sqrt{3}}{2}\right)^{n-1} \\
w_n(2) &= w_n(3) = \frac{\sqrt{3} - 5}{2\sqrt{3}} \left(\frac{1 - \sqrt{3}}{2}\right)^n + \frac{\sqrt{3} + 5}{2\sqrt{3}} \left(\frac{1 + \sqrt{3}}{2}\right)^n \\
w_n(4) &= w_n(5) = \frac{\sqrt{3} - 5}{2\sqrt{3}} \left(\frac{1 - \sqrt{3}}{2}\right)^{n-1} + \frac{\sqrt{3} + 5}{2\sqrt{3}} \left(\frac{1 + \sqrt{3}}{2}\right)^{n-1}
\end{align*}
\]

(d) The generating function \(W\) is given by

\[
\begin{align*}
W(1, t) &= 1 + \frac{\sqrt{3} - 5}{\sqrt{3}} \frac{t(1 - \sqrt{3})}{2 - t(1 - \sqrt{3})} + \frac{\sqrt{3} + 5}{\sqrt{3}} \frac{t(1 + \sqrt{3})}{2 - t(1 + \sqrt{3})}, \quad |t| < \frac{2}{1 + \sqrt{3}} \\
W(2, t) &= W(3, t) = \frac{\sqrt{3} - 5}{2\sqrt{3}} \frac{1}{2 - t(1 - \sqrt{3})} + \frac{\sqrt{3} + 5}{2\sqrt{3}} \frac{1}{2 - t(1 + \sqrt{3})}, \quad |t| < \frac{2}{1 + \sqrt{3}} \\
W(4, t) &= W(5, t) = 1 + \frac{\sqrt{3} - 5}{2\sqrt{3}} \frac{t(1 - \sqrt{3})}{2 - t(1 - \sqrt{3})} + \frac{\sqrt{3} + 5}{2\sqrt{3}} \frac{t(1 + \sqrt{3})}{2 - t(1 + \sqrt{3})}, \quad |t| < \frac{2}{1 + \sqrt{3}}
\end{align*}
\]

Exercise 2.5 For the directed diamond graph \((S, \rightarrow)\):

(a) The adjacency matrix is

\[
L = \begin{bmatrix}
0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0
\end{bmatrix}
\]

(b) Eigenvalue \(2^{1/3}\) has eigenvector \((2^{1/3}, 2^{-1/3}, 2^{-1/3}, 1)\). Eigenvalue 0 has eigenvector \((0, -1, 1, 0)\). There is also a pair of conjugate complex eigenvalues: \(2^{-2/3}(-1 \pm \sqrt{3}i)\).

(c) The left walk function of order \(n \in \mathbb{N}\) is

\[
\begin{align*}
w_n &= \left(2^{n/3}, 2^{n/3}, 2^{n/3}, 2^{n/3}\right), \quad n = 0 \mod 3 \\
w_n &= \left(2^{(n-1)/3}, 2^{(n-1)/3}, 2^{(n-1)/3}, 2^{(n+2)/3}\right), \quad n = 1 \mod 3 \\
w_n &= \left(2^{(n+1)/3}, 2^{(n-2)/3}, 2^{(n-2)/3}, 2^{(n+1)/3}\right), \quad n = 2 \mod 3
\end{align*}
\]
(d) The generating function $\Gamma$ is given by

$$W(1, t) = \frac{1 + t + 2t^2}{1 - 2t^3}, \quad |t| < \frac{1}{2^{1/3}}$$

$$W(2, t) = W(3, t) = \frac{1 + t + t^2}{1 - 2t^3}, \quad |t| < \frac{1}{2^{1/3}}$$

$$W(4, t) = \frac{1 + 2t + 2t^2}{1 - 2t^3}, \quad |t| < \frac{1}{2^{1/3}}$$

Chapter 4

Exercise 4.1 For the diamond graph $(S, \leftrightarrow)$:

(a) The probability function $F$ of $X$ for $(S, \leftrightarrow)$ is given by

$$F(1) = F(3) = f(2) + f(4), \quad F(2) = f(1) + f(3) + f(4), \quad F(4) = f(1) + f(3) + f(2)$$

(b) The rate function $r$ of $X$ for $(S, \leftrightarrow)$ is given by

$$r(1) = \frac{f(1)}{f(2) + f(4)}, \quad r(2) = \frac{f(2)}{f(1) + f(3) + f(4)}$$

$$r(3) = \frac{f(3)}{f(2) + f(4)}, \quad r(4) = \frac{f(4)}{f(1) + f(2) + f(3)}$$

(c) No. The adjacency matrix $A$ has eigenvalue 0 and the corresponding eigenvector $(1, 0, -1, 0)$ sums to 0.

Exercise 4.2 For the bull graph $(S, \leftrightarrow)$:

(a) The probability function $F$ of $X$ for $(S, \leftrightarrow)$ is given by

$$F(1) = f(2) + f(3), \quad F(2) = f(1) + f(3) + f(4), \quad F(4) = f(2), \quad F(5) = f(3)$$

(b) The rate function $r$ of $X$ for $(S, \leftrightarrow)$ is given by

$$r(1) = \frac{f(2)}{f(3)}, \quad r(2) = \frac{f(2)}{f(1) + f(3) + f(4)}$$

$$r(3) = \frac{f(3)}{f(1) + f(2) + f(5)}, \quad r(4) = \frac{f(4)}{f(2)}, \quad r(5) = \frac{f(5)}{f(3)}$$

(c) Yes. The adjacency matrix $\delta$ has eigenvalue 0 and so is not invertible but the corresponding eigenvector $(-1, 0, 0, 1, 1)$ does not sum to 0. Specifically, we can recover $f$ from $F$ by

$$f(1) = F(2) + F(3) - 1, \quad f(2) = F(4), \quad f(3) = F(5)$$

$$f(4) = 1 - F(3) - F(5), \quad f(5) = 1 - F(2) - F(4)$$

Exercise 4.3 For the directed diamond graph $(S, \rightarrow)$:

(a) The probability function $F$ of $X$ for $(S, \rightarrow)$ is given by

$$F(1) = f(2) + f(3), \quad F(2) = F(3) = f(4), \quad F(4) = f(1)$$

(b) The rate function $r$ of $X$ for $(S, \rightarrow)$ is given by

$$r(1) = \frac{f(2)}{f(2) + f(3)}, \quad r(2) = \frac{f(2)}{f(4)}, \quad r(3) = \frac{f(3)}{f(4)}, \quad r(4) = \frac{f(4)}{f(1)}$$

(c) No. The adjacency matrix $\Delta$ has eigenvalue 0 and the corresponding eigenvector sums to 0. More specifically, $F$ determines $f(1), f(4)$ and $f(2) + f(3)$, but not $f(2)$ and $f(3)$ individually.
Exercise 4.4  Note that here is a loop at \( x \) and this is the only edge from \( x \). There is a loop at \( y \) and at least one other edge from \( y \) (to \( w \)). Suppose \( f \) is a probability density function on \( S \) with constant rate, and let \( F \) denote the right probability function of \( f \) for \((S, \rightarrow)\). Then \( F(x) = f(x) \) so the rate constant must be 1. But then \( f(y) = F(y) \) but also \( F(y) \geq f(y) + f(w) \). So we have \( f(w) = F(w) = 0 \) and this contradicts the support assumption that \( F(t) > 0 \) for all \( t \in S \).

Exercise 4.5  For the diamond graph \((S, \leftrightarrow)\),

(a) The rate constant is \( \alpha = 2/(1 + \sqrt{17}) \), the reciprocal of the largest eigenvalue.

(b) The density function \( f \) and probability function \( F \) are given by

\[
 f(1) = f(3) = \frac{5 - \sqrt{17}}{4}, \quad f(2) = f(4) = \frac{\sqrt{17} - 3}{4}
\]

\[
 F(1) = F(3) = \frac{\sqrt{17} - 3}{2}, \quad F(2) = F(4) = \frac{7 - \sqrt{17}}{4}
\]

(c) The density function \( f_n \) is given by

\[
 f_n(1) = f_n(3) = \left(1 - \frac{4}{\sqrt{17}}\right) \left(\frac{1 - \sqrt{17}}{1 + \sqrt{17}}\right)^{n-1} + \frac{1}{4} \left(1 - \frac{1}{\sqrt{17}}\right)
\]

\[
 f_n(3) = f_n(4) = -\left(1 - \frac{4}{\sqrt{17}}\right) \left(\frac{1 - \sqrt{17}}{1 + \sqrt{17}}\right)^{n-1} + \frac{1}{4} \left(1 + \frac{1}{\sqrt{17}}\right)
\]

(d) The limiting values are

\[
 \lim_{n \to \infty} f_n(1) = \lim_{n \to \infty} f_n(3) = \frac{1}{4} \left(1 - \frac{1}{\sqrt{17}}\right)
\]

\[
 \lim_{n \to \infty} f_n(2) = \lim_{n \to \infty} f_n(4) = \frac{1}{4} \left(1 + \frac{1}{\sqrt{17}}\right)
\]

Moreover, \( \lim_{n \to \infty} f_n \) is \( fF \) normalized and hence is the invariant density function.

Exercise 4.6  For the bull graph \((S, \leftrightarrow)\),

(a) The rate constant is \( \alpha = 2/(1 + \sqrt{13}) \), the reciprocal of the largest eigenvalue.

(b) The density function \( f \) and probability function \( F \) are given by

\[
 f(1) = \frac{5 - \sqrt{13}}{6}, \quad f(2) = f(3) = \frac{\sqrt{13} - 2}{6}, \quad f(4) = f(5) = \frac{5 - \sqrt{13}}{12}
\]

\[
 F(1) = \frac{\sqrt{13} - 2}{3}, \quad F(2) = F(3) = \frac{11 - \sqrt{13}}{12}, \quad F(4) = F(5) = \frac{\sqrt{13} - 2}{6}
\]

(c) The density function \( f_n \) is given by

\[
 f_n(1) = f_n(3) = \left(\frac{14}{3\sqrt{13}} - \frac{4}{3}\right) \left(\frac{1 - \sqrt{13}}{1 + \sqrt{13}}\right)^{n-2} + \frac{1}{3} \left(1 - \frac{1}{\sqrt{13}}\right)
\]

\[
 f_n(2) = f_n(3) = \left(1 - \frac{7}{2\sqrt{13}}\right) \left(\frac{1 - \sqrt{13}}{1 + \sqrt{13}}\right)^{n-2} + \frac{1}{4} \left(1 + \frac{1}{\sqrt{13}}\right)
\]

\[
 f_n(4) = f_n(5) = \left(-\frac{1}{3} + \frac{7}{6\sqrt{13}}\right) \left(\frac{1 - \sqrt{13}}{1 + \sqrt{13}}\right)^{n-2} + \frac{1}{12} \left(1 - \frac{1}{\sqrt{13}}\right)
\]
(d) The limiting values are

\[
\lim_{n \to \infty} f_n(1) = \frac{1}{3} \left( 1 - \frac{1}{\sqrt{13}} \right)
\]

\[
\lim_{n \to \infty} f_n(2) = \lim_{n \to \infty} f_n(3) = \frac{1}{4} \left( 1 + \frac{1}{\sqrt{13}} \right)
\]

\[
\lim_{n \to \infty} f_n(4) = \lim_{n \to \infty} f_n(5) = \frac{1}{12} \left( 1 - \frac{1}{\sqrt{13}} \right)
\]

Moreover, \( \lim_{n \to \infty} f_n \) is \( fF \) normalized and hence is the invariant density function.

**Exercise 4.7** For the directed diamond graph \((S, \to)\),

(a) The rate constant is \( \alpha = 2^{-1/3} \), the reciprocal of the largest eigenvalue.

(b) The density function \( f \) and the right probability function \( F \) for the distribution with constant rate are given by

\[
f(1) = \frac{2^{1/3}}{1 + 2^{1/3} + 2^{2/3}}, \quad f(2) = f(3) = \frac{2^{-1/3}}{1 + 2^{1/3} + 2^{2/3}}, \quad f(4) = \frac{1}{1 + 2^{1/3} + 2^{2/3}}
\]

\[
F(1) = \frac{2^{2/3}}{1 + 2^{1/3} + 2^{2/3}}, \quad F(2) = f(3) = \frac{1}{1 + 2^{1/3} + 2^{2/3}}, \quad F(4) = \frac{2^{1/3}}{1 + 2^{1/3} + 2^{2/3}}
\]

(c) The density function \( f_n \) of \( X_n \) is given as follows: If \( n = 0 \mod 3 \) then \( f_n = f \). If \( n = 1 \mod 3 \) then \( f_n(x) = 2^{-1/3}f(x) \) for \( x \in \{1, 2, 3\} \) and \( f_n(4) = 2^{2/3}f(4) \). If \( n = 2 \mod 3 \) then \( f_n(x) = 2^{1/3}f(x) \) for \( x \in \{1, 4\} \) and \( f_n(x) = 2^{-2/3}f(x) \) for \( x \in \{2, 3\} \). Note that the random walk \( X \) is periodic with period 3.

**Chapter 5**

**Exercise 5.1**

(a) \( X \) has probability function \( F \) for \((0, \infty), +)\) given by

\[
F(x) = \frac{1}{(x+1)^a}, \quad 0 \leq x < \infty
\]

(b) The transition density \( P \) of the random walk on \((0, \infty), \leq)\) associated with \( X \) is

\[
P(x,y) = \frac{a(x+1)^a}{(y+1)^{a+1}}, \quad 0 \leq x \leq y < \infty
\]

(c) The transition density \( Q \) of the random walk on \((0, \infty), +)\) associated with \( X \) is

\[
Q(x,y) = \frac{a}{(y-x+1)^{a+1}}, \quad 0 \leq x \leq y < \infty
\]

**Chapter 6**

**Exercise 6.1** For the reflexive completion \((S, \leftrightarrow)\) of the diamond graph \((S, \to)\),

(a) The walk function of order \( n \in \mathbb{N} \) is given by

\[
\bar{\gamma}_n(1) = \bar{\gamma}_n(3) = \frac{1}{\sqrt{17}} \left[ \left( \frac{3 + \sqrt{17}}{2} \right)^{n+1} - \left( \frac{3 - \sqrt{17}}{2} \right)^{n+1} \right]
\]

\[
\bar{\gamma}_n(2) = \bar{\gamma}_n(4) = \frac{\sqrt{17} - 5}{2\sqrt{17}} \left( \frac{3 - \sqrt{17}}{2} \right)^n + \frac{\sqrt{17} + 5}{2\sqrt{17}} \left( \frac{3 + \sqrt{17}}{2} \right)^n
\]
(b) The rate constant is $\beta = 2/(3 + \sqrt{17})$. The constant rate distribution is the same as for the regular diamond graph, with density function $f$ given by

$$f(1) = f(3) = \frac{5 - \sqrt{17}}{4}, \quad f(2) = f(4) = \frac{\sqrt{17} - 3}{4}$$

### Chapter 8

#### Exercise 7.1
(a) The adjacency operator $A^n$ of order $n \in \mathbb{N}$ for $(S, \equiv)$ is given by

$$A^n u(x) = \beta^{n-1} \int_S u(y) d\lambda(y), \quad x \in S$$

(b) The walk function $\gamma_n$ of order $n \in \mathbb{N}$ for $(S, \equiv)$ is constant on $S$: $\gamma_n = \beta^n$.

(c) The generating function $\Gamma$ for $(S, \equiv)$ is given by $\Gamma(x, t) = 1/(1 - \beta t)$ for $x \in S$ and $t \in (-1, 1)$.

(d) The probability function $F$ of a random variable $X$ for $(S, \equiv)$ is constant on $S$: $F = 1$. If $X$ has density $f$ then the rate function of $X$ for $(S, \equiv)$ is $r = f$.

(e) The only distribution with constant rate for $(S, \equiv)$ is the uniform distribution on $S$, with rate $1/\beta$.

(f) The random walk $X = (X_1, X_2, \ldots)$ on $(S, \equiv)$ associated with a density function $f$ is a sequence of independent variables, each with density function $f$.

#### Exercise 7.2
(a) The walk function $\gamma_n$ of order $n \in \mathbb{N}$ for $(S, \equiv)$ is constant on $S$: $\gamma_n = 1$.

(b) The generating function $\Gamma$ for $(S, \equiv)$ is given by $\Gamma(x, t) = 1/(1 - t)$ for $x \in S$ and $t \in (-1, 1)$.

(c) If $X$ is a random variable with values in $S$ and density function $f$ then the probability function $F$ of $X$ for $(S, \equiv)$ is $F = f$. So every distribution on $S$ has constant rate $1$ for $(S, \equiv)$.

(d) The random walk on $(S, \equiv)$ associated with the distribution of a random variable $X$ has the form $(X, X, \ldots)$.

#### Exercise 7.3
The induced equivalence relation $\equiv$ corresponds to the discrete graph $(\mathbb{N}, =)$ and the index function $\varphi$ is given by $\varphi(x) = \lfloor x \rfloor$ for $x \in [0, \infty)$.

(a) The walk function $\gamma_n$ on $(S, \equiv)$ of order $n \in \mathbb{N}$ is constant on $[0, \infty)$: $\gamma_n = 1$.

(b) The generating function $\Gamma$ on $(S, \equiv)$ is given by $\Gamma(x, t) = 1/(1 - t)$ for $x \in [0, \infty)$ and $t \in (-1, 1)$.

(c) Suppose that $\hat{f}$ is a probability density function on $\mathbb{N}$. Define $f(x) = \hat{f}(k)$ for $x \in [k, k + 1)$. Then $f$ is a density on $[0, \infty)$ which has constant rate $1$ for $([0, \infty), \equiv)$.

#### Exercise 7.4
(a) The walk function $\gamma_n$ of order $n \in \mathbb{N}$ for $(S, \leftrightarrow)$ is given by

$$\gamma_n(x) = \lfloor (k-1)\beta \rfloor^n, \quad x \in S$$

(b) The generating function $\Gamma$ for $(S, \leftrightarrow)$ is given by

$$\Gamma(x, t) = \frac{1}{1 - (k-1)\beta t}, \quad x \in S, \quad |t| < \frac{1}{(k-1)\beta}$$

(c) The uniform distribution on $S$ is the unique constant rate distribution for $(S, \leftrightarrow)$, with rate $\alpha = 1/(k-1)\beta$. 

Chapter 11

Exercise 11.1  Left walk function $w_n$ of order $n \in \mathbb{N}$.

(a) For the graph $(\mathbb{N}, <)$, $w_n(x) = \binom{x}{n}$ for $x \in \mathbb{N}$. Note that this expression is correct when $n = 0$. Assume that the expression is correct for a given $n \in \mathbb{N}$. Then

$$w_{n+1}(x) = \sum_{z=0}^{x-1} w_n(z) = \sum_{z=0}^{x-1} \binom{z}{n} = \binom{x}{n+1}$$

by a binomial identity. For the combinatorial proof, to construct a walk of length $n$ in $(\mathbb{N}, <)$ ending in $x$ we need to select a subset of size $n$ from $\{0, \ldots, x-1\}$. The number of ways to do this is $\binom{x}{n}$.

(b) For the graph $(\mathbb{N}, \uparrow)$, $w_n(x) = 1 (x \geq n)$ for $x \in \mathbb{N}$. A walk is a sequence of consecutive nonnegative integers. So if $x < n$ there are no walks of length $n$ ending in $x$. If $x \geq n$ there is one such walk, namely $(x-n, x-n+1, \ldots, x-1, x)$.

(c) For the graph $(\mathbb{N}, \Rightarrow)$, $w_n(x) = \sum_{k=0}^{x} \binom{n}{k}$ for $x \in \mathbb{N}$. This follows from general results on reflexive completion. Note that $w_n(x) = 2^n$ if $x \geq n$.

Exercise 11.2  Left generating functions:

(a) For the graph $(\mathbb{N}, <)$,

$$W(x,t) = \sum_{n=0}^{\infty} w_n(x) t^n = \sum_{n=0}^{x} \binom{x}{n} t^n = (1 + t)^x, \quad x \in \mathbb{N}, \; t \in \mathbb{R}$$

(b) For the graph $(\mathbb{N}, \uparrow)$,

$$W(x,t) = \sum_{n=0}^{\infty} w_n(x) t^n = \sum_{n=0}^{x} t^n = \frac{1 - t^{x+1}}{1 - t}$$

It’s understood that the fraction is $x+1$ if $t = 1$.

(c) For the graph $(\mathbb{N}, \Rightarrow)$,

$$W(x,t) = \sum_{n=0}^{\infty} w_n(x) t^n = \sum_{n=0}^{x} \sum_{k=0}^{n} \binom{n}{k} t^n = \sum_{n=0}^{x} \sum_{k=0}^{\infty} \binom{n}{k} t^n$$

$$= \sum_{k=0}^{x} \frac{1}{1-t} \left( \frac{t}{1-t} \right)^k = \frac{1 - \left( \frac{t}{1-t} \right)^{x+1}}{1 - \frac{t}{1-t}}$$

$$= \frac{t^{x+1} - (1-t)^{x+1}}{(1-t)^{x+1}(2t-1)}, \quad x \in \mathbb{N}, \; |t| < 1$$

Exercise 11.3

(a) For the graph $(\mathbb{N}, <)$,

$$F(x) = \mathbb{P}(x < X) = \sum_{y=x+1}^{\infty} f(y), \quad x \in \mathbb{N}$$

$$r(x) = \frac{f(x)}{F(x)} = \frac{f(x)}{\sum_{y=x+1}^{\infty} f(y)}, \quad x \in \mathbb{N}$$

The density function $f$ can be recovered from $F$ by $f(0) = 1 - F(0)$ and $f(x) = F(x-1) - F(x)$ for $x \in \mathbb{N}_+$. 
(b) For the graph \((\mathbb{N}, \uparrow)\),

\[
F(x) = \mathbb{P}(x \uparrow X) = f(x + 1), \quad x \in \mathbb{N}
\]

\[
r(x) = \frac{f(x)}{F(x)} = \frac{f(x)}{f(x + 1)}, \quad x \in \mathbb{N}
\]

The density function \(f\) can be recovered from \(F\) by \(f(x) = F(x) - F(x - 1)\) for \(x \in \mathbb{N}_+\) and then \(f(0) = 1 - \sum_{x=1}^{\infty} f(x)\).

(c) For the graph \((\mathbb{N}, \uparrow)\),

\[
F(x) = \mathbb{P}(x \uparrow X) = f(x) + f(x + 1), \quad x \in \mathbb{N}
\]

\[
r(x) = \frac{f(x)}{F(x)} = \frac{f(x)}{f(x) + f(x + 1)}, \quad x \in \mathbb{N}
\]

The density function \(f\) can be recovered from \(F\) by

\[
f(x) = \sum_{n=0}^{x-1} (-1)^{n-x} F(n) + (-1)^x \left[ 2 - \sum_{n=0}^{\infty} F(n) \right], \quad x \in \mathbb{N}
\]

**Exercise 11.4**  For a graph on \(\mathbb{N}\) and a random variable \(X\) on \(\mathbb{N}\), the standard moment result is

\[
\sum_{x=0}^{\infty} w_n(x) F(x) = \mathbb{E}[w_{n+1}(X)], \quad n \in \mathbb{N}
\]

where \(F\) is the right probability function of \(X\) and \(w_n\) is the walk function of order \(n \in \mathbb{N}\) for the graph.

(a) For the graph \((\mathbb{N}, <)\),

\[
\sum_{x=1}^{\infty} \binom{x}{n} \mathbb{P}(X > x) = \mathbb{E}\left[ \binom{X}{n+1} \right], \quad n \in \mathbb{N}
\]

(b) For the graph \((\mathbb{N}, \uparrow)\),

\[
\sum_{x=0}^{\infty} 1(x \geq n) \mathbb{P}(X = x + 1) = \mathbb{E}[1(X \geq n + 1)], \quad n \in \mathbb{N}
\]

Note that this is equivalent to the obvious result

\[
\sum_{x=n}^{\infty} \mathbb{P}(X = x + 1) = \mathbb{P}(X \geq n + 1), \quad n \in \mathbb{N}
\]

(c) For the graph \((\mathbb{N}, \uparrow)\),

\[
\sum_{x=0}^{\infty} \sum_{k=0}^{x} \binom{n}{k} [\mathbb{P}(X = x) + \mathbb{P}(X = x + 1)] = \mathbb{E}\left[ \sum_{k=0}^{X} \binom{n+1}{k} \right], \quad n \in \mathbb{N}
\]

When \(n = 0\), this reduces to the obvious result

\[
\sum_{x=0}^{\infty} [\mathbb{P}(X = x) + \mathbb{P}(X = x + 1)] = 1 + \mathbb{P}(X \geq 1)
\]
Exercise 11.5  Let $F$ denote the right probability function of $X$ for $(\mathbb{N},<)$ and recall that $f(0) = 1 - F(0)$ and $f(x) = F(x) - F(x)$ for $x \in \mathbb{N}$. For $\alpha \in (0, \infty)$, the constant rate property $f = \alpha F$ has the unique solution

$$
F(x) = \left( \frac{1}{1+\alpha} \right)^{x+1}, \quad x \in \mathbb{N}
$$

$$
f(x) = \frac{\alpha}{1+\alpha} \left( \frac{1}{1+\alpha} \right)^x, \quad x \in \mathbb{N}
$$

Let $Y = X + 1$, so that $Y$ has values in $\mathbb{N}_+$. Then $Y$ has density function $g$ given by $g(y) = f(y-1)$ for $y \in \mathbb{N}_+$, and the right probability function $G$ of $Y$ for $(\mathbb{N}_+,<)$ is given by $G(y) = F(y-1)$ for $y \in \mathbb{N}_+$. If $X$ has the distribution with constant rate $\alpha \in (0, \infty)$ for $(\mathbb{N},<)$ above, then $Y$ is memoryless for $(\mathbb{N}_+,+)$ and has constant rate $\alpha$ for the associated graph $(\mathbb{N},<)$. Hence $Y$ is exponential for $(\mathbb{N}_+,+)$. Conversely, if $Y$ is memoryless for $(\mathbb{N}_+,+)$ then $G(y) = (1-\beta)^y$ for $y \in \mathbb{N}_+$ and hence $g(y) = \beta(1-\beta)^{y-1}$ for $y \in \mathbb{N}_+$ where $\beta = G(1) \in (0,1)$. Hence $Y$ has constant rate $\alpha = \beta/(1-\beta) \in (0, \infty)$ for $(\mathbb{N}_+,<)$ and therefore $Y$ is exponential for $(\mathbb{N}_+,+)$. Then also $X = Y - 1$ has constant rate $\alpha$ for $(\mathbb{N},<)$.

Exercise 11.6  Let $f$ denote the density function of $X$ and let $F$ denote the right probability function of $X$ for $(\mathbb{N},\uparrow)$. Then $F(x) = f(x+1)$ for $x \in \mathbb{N}$, so the constant rate property is $f(x) = \alpha f(x+1)$ for $x \in \mathbb{N}$. Hence $f(x) = (1/\alpha)^x f(0)$ for $x \in \mathbb{N}$. In order for $f$ to be a proper density function, we must have $\alpha > 1$, in which case $f(x) = (1-1/\alpha)(1/\alpha)^x$ for $x \in \mathbb{N}$. The total order graph $(\mathbb{N},\leq)$ is uniform, so from our general theory, $X$ has constant rate $\alpha^n$ for $(\mathbb{N},\uparrow)$ for each $n \in \mathbb{N}$.

Exercise 11.7  The proof follows from general results on reflexive completion: $X$ has constant rate $\alpha$ for $(\mathbb{N},\dagger)$ if and only if $X$ has constant rate $\alpha/(1-\alpha)$ for $(\mathbb{N},\uparrow)$.

Exercise 11.8  Suppose that $Y = (Y_1, Y_2, \ldots)$ is the random walk on $(\mathbb{N},<)$ corresponding to the distribution with constant rate $\alpha \in (0, \infty)$.

(a) $Y$ has transition density $P$ given by

$$
P(x,y) = \alpha \left( \frac{1}{1+\alpha} \right)^{y-x}, \quad x,y \in \mathbb{N}, \ x < y
$$

(b) For $n \in \mathbb{N}_+$, $(Y_1, Y_2, \ldots, Y_n)$ has density function $g_n$ defined by

$$
g_n(y_1, y_2, \ldots, y_n) = \alpha^n \left( \frac{1}{1+\alpha} \right)^{y_{n+1}}, \quad (y_1, y_2, \ldots, y_n) \in \mathbb{N}^n, \ y_1 < y_2 < \cdots < y_n
$$

(c) For $n \in \mathbb{N}_+$, $Y_n$ has density function $f_n$ defined by

$$
f_n(y) = \alpha^n \binom{y}{n-1} \left( \frac{1}{1+\alpha} \right)^{y+1}, \quad y \in \{n-1, n, \ldots \}
$$

(d) Given $Y_{n+1} = y \in \{n, n+1, \ldots\}$, $(Y_1, Y_2, \ldots, Y_n)$ is uniformly distributed on the $(y-1)$ points in the set

$$
\{(y_1, y_2, \ldots, y_n) \in \mathbb{N}^n : y_1 < y_2 < \cdots < y_n < y\}
$$

The distribution of $Y_n$ in (c) is also closely related to the negative binomial distribution.

Exercise 11.9  Suppose that $Y = (Y_1, Y_2, \ldots)$ is the random walk on $(\mathbb{N},\uparrow)$ corresponding to the distribution with constant rate $\alpha \in (1, \infty)$.

(a) $Y$ has transition density $P$ given by

$$
P(x,x+1) = 1, \quad x \in \mathbb{N}
$$
(b) For $n \in \mathbb{N}_+$, $(Y_1, Y_2, \ldots, Y_n)$ has density function $g_n$ defined by

$$g_n(y_1, y_2, \ldots, y_n) = \alpha^{n-1} \frac{2\alpha - 1}{\alpha} \left( \frac{1 - \alpha}{\alpha} \right)^{y_n}$$

For $(y_1, y_2, \ldots, y_n) \in \mathbb{N}^n$ with $y_{k+1} \in \{y_k, y_k + 1\}$ for each $k \in \{1, 2, \ldots, n-1\}$.

(c) For $n \in \mathbb{N}_+$, $Y_n$ has density function $f_n$ defined by

$$f_n(y) = \frac{\alpha - 1}{\alpha} \left( \frac{1}{\alpha} \right)^{y-n+1}, \quad y \in \{n-1, n, n+1, \ldots\}$$

(d) Given $Y_{n+1} = y \in \{n, n+1, \ldots\}$, $(Y_1, Y_2, \ldots, Y_n) = (y-n, y-n-1, \ldots, y-1)$ with probability 1. These results are trivial. In part (b), the expression on the right is just the original exponential distribution at the starting point $y_1$. The density in part (c) is the original exponential distribution, shifted to the right by $n-1$. In part (d), if we know the value of $Y_{n+1}$, the previous variables are deterministic.

**Exercise 11.10** Suppose that $Y = (Y_1, Y_2, \ldots)$ is the random walk on $(\mathbb{N}, \uparrow)$ corresponding to the distribution with constant rate $\alpha \in (1/2, 1)$.

(a) $Y$ has transition density $P$ given by

$$P(x, x) = \alpha, \quad P(x, x+1) = 1 - \alpha; \quad x \in \mathbb{N}$$

(b) For $n \in \mathbb{N}_+$, $(Y_1, Y_2, \ldots, Y_n)$ has density function $g_n$ defined by

$$g_n(y_1, y_2, \ldots, y_n) = \alpha^{n-1} \frac{2\alpha - 1}{\alpha} \left( \frac{1 - \alpha}{\alpha} \right)^{y_n}$$

For $(y_1, y_2, \ldots, y_n) \in \mathbb{N}^n$ with $y_{k+1} \in \{y_k, y_k + 1\}$ for each $k \in \{1, 2, \ldots, n-1\}$.

(c) For $n \in \mathbb{N}_+$, $Y_n$ has density function $f_n$ defined by

$$f_n(y) = \frac{\alpha - 1}{\alpha} \left( \frac{1}{\alpha} \right)^{y-n+1}, \quad y \in \{n-1, n, n+1, \ldots\}$$

(d) Given $Y_{n+1} = y \in \mathbb{N}$, $(Y_1, Y_2, \ldots, Y_n)$ is uniformly distributed on the $\sum_{k=0}^{n-1} \binom{n}{k}$ points in the set

$\{(y_1, y_2, \ldots, y_n) \in \mathbb{N}^n : y_{k+1} \in \{y_k, y_k + 1\} \text{ for each } k \in \{1, 2, \ldots, n-1\} \text{ and } y_n \in \{y-1, y\}\}$

**Exercise 11.11** Recall that if $X$ is the random walk corresponding to a distribution of a random variable $X$ with constant rate $\alpha \in (0, \infty)$ then

$$\mathbb{E}(N_x) = \mathbb{E}[W(X, \alpha); X \leq x]$$

where $W$ is the left generating function for the graph.

(a) For the graph $(\mathbb{N}, <)$,

$$\mathbb{E}(N_x) = \mathbb{E} \left[(1 + \alpha)^X; X \leq x\right] = \sum_{t=0}^{x} (1 + \alpha)^t \frac{\alpha}{1 + \alpha} \left( \frac{1}{1 + \alpha} \right)^t = \frac{\alpha}{1 + \alpha}(x + 1), \quad x \in \mathbb{N}$$

(b) For the graph $(\mathbb{N}, \uparrow)$,

$$\mathbb{E}(N_x) = \mathbb{E} \left[\frac{\alpha^{x+1} - 1}{\alpha - 1}; X \leq x\right] = \sum_{t=0}^{x} \frac{\alpha^{t+1} - 1}{\alpha - 1} \frac{1}{\alpha} \left( \frac{1}{\alpha} \right)^t = \sum_{t=0}^{x} \frac{\alpha^{t+1} - 1}{\alpha^{t+1}}$$

$$= \sum_{t=0}^{x} \left[1 - \left(\frac{1}{\alpha}\right)^{t+1}\right] = (x + 1) - \frac{\alpha^{x+1} - 1}{\alpha^{x+2} - \alpha^{x+1}}, \quad x \in \mathbb{N}$$
(c) For the graph \((\mathbb{N}, \uparrow)\),

\[
E(N_x) = \mathbb{E}\left[\frac{\alpha^{x+1} - (1 - \alpha)^{x+1}}{(1 - \alpha)^{x+1}(2\alpha - 1)}; X \leq x\right] = \sum_{t=0}^{x} \frac{\alpha^{t+1} - (1 - \alpha)^{t+1}}{(1 - \alpha)^{t+1}(2\alpha - 1)} \frac{2\alpha - 1}{\alpha} \left(\frac{1 - \alpha}{\alpha}\right)^t
\]

\[
= \sum_{t=0}^{x} \frac{\alpha^{t+1} - (1 - \alpha)^{t+1}}{(1 - \alpha)^{t+1}} = \frac{1}{1 - \alpha} \sum_{t=0}^{x} \left[1 - \left(\frac{1 - \alpha}{\alpha}\right)^{t+1}\right]
\]

\[
= \frac{1}{1 - \alpha} (x + 1) - \left[1 - \left(\frac{1 - \alpha}{\alpha}\right)^{x+1}\right], \quad x \in \mathbb{N}
\]

**Exercise 11.12** Recall that in general, if a random walk corresponds to a distribution with constant rate \(\alpha \in (0, \infty)\) for a graph on \(\mathbb{N}\), then the density function \(g\) of \(Y_N\) is given by

\[
g(x) = qaW[x,(1 - q)\alpha]F(x), \quad x \in S
\]

where \(W\) is the left generating function and \(F\) is the right probability function of the distribution for the graph.

(a) For the graph \((\mathbb{N}, \prec)\), the left generating function \(\Gamma\) is given by \(\Gamma(x,t) = (1 + t)^x\) for \(x \in \mathbb{N}\) and \(t \in \mathbb{R}\), and the right probability function \(F\) is given by \(F(x) = 1/(1 + \alpha)^{x+1}\) for \(x \in \mathbb{N}\). Hence

\[
g(x) = \frac{\alpha}{1 + \alpha} \left(\frac{1 + \alpha - aq}{1 + \alpha}\right)^x, \quad x \in \mathbb{N}
\]

(b) For the graph \((\mathbb{N}, \uparrow)\), the left generating function \(\Gamma\) is given by \(W(x,t) = (t^{x+1} - 1)/(t - 1)\) for \(x \in \mathbb{N}\) and \(t \in \mathbb{R}\), and the right probability function \(F\) is given by \(F(x) = (\alpha - 1)/\alpha^{x+2}\) for \(x \in \mathbb{N}\). Hence

\[
g(x) = \frac{(\alpha - 1)q(\alpha(1 - q)]^{x+1} - 1}{\alpha(1 - q) - 1} \left(\frac{1}{\alpha}\right)^x, \quad x \in \mathbb{N}
\]

(c) For the graph \((\mathbb{N}, \downarrow)\), the left generating function \(W\) is given by

\[
W(x,t) = \frac{t^{x+1} - (1 - t)^{x+1}}{(1 - t)^{x+1}(2t - 1)}, \quad x \in \mathbb{N}, t \in (-1, 1)
\]

and the right probability function \(F\) is given by \(F(x) = (2\alpha - 1)(1 - \alpha)^x/\alpha^{x+2}\) for \(x \in \mathbb{N}\). Hence

\[
g(x) = \frac{q(2\alpha - 1)[\alpha(1 - q)]^{x+1} - [1 - \alpha(1 - q)]^{x+1}}{[\alpha(1 - q) - 1][2\alpha(1 - q) - 1]} \left(\frac{1 - \alpha}{\alpha}\right)^x, \quad x \in \mathbb{N}
\]

**Chapter 12**

**Exercise 12.1** Left walk function \(\gamma_n\) of order \(n \in \mathbb{N}\)

(a) For the graph \((S, \uparrow)\),

\[
\gamma_n(x) = 1[d(x) \geq n], \quad x \in S
\]

For \(x \in S\), there is a single walk in \((S, \uparrow)\) of length \(n\) ending in \(x\) if \(n \leq d(x)\). If \(n > d(x)\) there is no walk ending in \(x\).

(b) For the graph \((S, \downarrow)\),

\[
\gamma_n(x) = \sum_{k=0}^{d(x)} \binom{n}{k}, \quad x \in S
\]

This follows from (a) and basic results on reflexive completion. Note that \(\gamma_n(x) = 2^n\) if \(n \leq d(x)\).

(c) For the graph, \((S, \prec)\),

\[
\gamma_n(x) = \binom{d(x)}{n}, \quad x \in S
\]

If \(n \leq d(x)\), then to construct a walk of length \(n\) ending in \(x\) in the graph \((S, \prec)\), we need to select a sample of size \(n\) from the first \(d(x)\) vertices on the unique path of length \(d(x)\) in \((S, \uparrow)\) from \(e\) to \(x\). This can also be obtained from Proposition 12.1 using basic results on reflexive completion.
Exercise 12.2  Left generating function $\Gamma$

(a) For the graph $(S, \uparrow)$,

$$\Gamma(x, t) = \sum_{n=0}^{\infty} 1[d(x) \geq n] t^n = \sum_{n=0}^{d(x)} t^n = \frac{1 - t^{d(x)+1}}{1-t}, \quad x \in S, \ |t| < 1$$

It’s understood that the fraction is $d(x) + 1$ if $t = 1$.

(b) For the graph $(S, \upharpoonright)$,

$$\Gamma(x, t) = \sum_{n=0}^{\infty} \sum_{k=0}^{d(x)} \binom{n}{k} t^n = \sum_{k=0}^{d(x)} \sum_{n=k}^{\infty} \binom{n}{k} t^n$$

$$= \sum_{k=0}^{d(x)} \frac{1}{1-t} \left( \frac{t}{1-t} \right)^k = \frac{1 - \left( \frac{t}{1-t} \right)^{d(x)+1}}{1 - \frac{t}{1-t}}$$

$$= \frac{t^{d(x)+1} - (1-t)^{d(x)+1}}{(1-t)^{d(x)+1}(2t-1)}, \quad x \in S, \ |t| < 1$$

(c) For the graph $(S, \prec)$,

$$\Gamma(x, t) = \sum_{n=0}^{d(x)} \binom{d(x)}{n} t^n = (1+t)^{d(x)}, \quad x \in S, \ t \in \mathbb{R}$$

Exercise 12.2  Suppose that random variable $X$ takes values in $S$ and that $n \in \mathbb{N}$. From the standard moment result, recall that

$$\sum_{x \in S} \gamma_n(x) F(x) = \mathbb{E}[\gamma_{n+1}(X)], \quad n \in \mathbb{N}$$

where $\gamma_n$ is the left walk function of order $n \in \mathbb{N}$ for the graph and $F$ is the right probability function of $X$ for the graph.

(a) For the graph $(S, \uparrow)$, $\gamma_n(x) = 1[d(x) \geq n]$ and $F(x) = \mathbb{P}(x \uparrow X)$ for $x \in S$. Hence

$$\sum_{d(x) \geq n} \mathbb{P}(x \uparrow X) = \mathbb{P}(d(X) \geq n+1)$$

(b) For the graph $(S, \upharpoonright)$, $\gamma_n(x) = \sum_{k=0}^{n \choose k} d(x)$ and $F(x) = \mathbb{P}(x \upharpoonright X) = \mathbb{P}(x = X) + \mathbb{P}(x \upharpoonright X)$ for $x \in S$. Hence

$$\sum_{x \in S} \sum_{k=0}^{n \choose k} \mathbb{P}(x \upharpoonright X) = \mathbb{E} \left[ \sum_{k=0}^{d(x)} {n \choose k} \right]$$

(c) For the graph $(S, \prec)$, $\gamma_n(x) = \binom{d(x)}{n}$ and $F(x) = \mathbb{P}(X \succ x)$ for $x \in S$. Hence

$$\sum_{x \in S} \binom{d(x)}{n} \mathbb{P}(X \succ x) = \mathbb{E} \left[ \binom{d(X)}{n+1} \right]$$

Exercise 12.3  Recall that if $X$ is the random walk on a graph corresponding to a distribution with constant rate $\alpha \in (0, \infty)$ then $X_n$ has density function $f_n$ given by

$$f_n(x) = \alpha^{n-1} \gamma_{n-1}(x) f(x), \quad x \in S$$

where of course, $\gamma_{n-1}$ is the left walk function of order $n - 1$ for the graph, and where $f$ the density function of the underlying distribution.
(a) For the graph \((S, \uparrow)\), the rate constant is \(1/(1-\alpha)\) and \(\gamma_{n-1}(x) = 1[d(x) \geq n-1]\). Hence

\[
f_n(x) = \frac{1}{(1-\alpha)^{n-1}} f(x), \quad d(x) \geq n-1
\]

So in particular, \(f_n(x) = 0\) if \(d(x) < n-1\).

(b) For the graph \((S, \uparrow)\), the rate constant is \(1/(2-\alpha)\) and \(\gamma_{n-1}(x) = \sum_{k=0}^{d(x)} \binom{n-1}{k} \). Hence

\[
f_n(x) = \frac{1}{(2-\alpha)^{n-1}} \left[ \sum_{k=0}^{d(x)} \binom{n-1}{k} \right] f(x), \quad x \in S
\]

Note that if \(d(x) \geq n-1\) then

\[
f_n(x) = \left( \frac{2}{2-\alpha} \right)^{n-1} f(x)
\]

(c) For the graph \((S, \prec)\), the rate constant is \(\alpha/(1-\alpha)\) and \(\gamma_{n-1}(x) = \binom{d(x)}{n} \). Hence

\[
f_n(x) = \left( \frac{\alpha}{1-\alpha} \right)^{n-1} \binom{d(x)}{n-1} f(x), \quad x \in S
\]

Note that \(f_n(x) = 0\) if \(d(x) < n-1\).

**Chapter 13**

**Exercise 13.1** Let \(\gamma_n\) denote the left walk function of order \(n \in \mathbb{N}\). Then

(a) For the graph \((S, \uparrow)\),

\[
\gamma_n(x) = 1[d(x) \geq n], \quad x \in S
\]

(b) For the graph \((S, \uparrow)\),

\[
\gamma_n(x) = \sum_{k=0}^{d(x)} \binom{n}{k}, \quad x \in S
\]

(c) For the graph \((S, \prec)\),

\[
\gamma_n(x) = \binom{d(x)}{n}, \quad x \in S
\]

**Exercise 13.2** Left generating function \(\Gamma\)

(a) For the graph \((S, \uparrow)\),

\[
\Gamma(x, t) = \frac{1 - t^{d(x)+1}}{1-t}, \quad x \in S, \ |t| < 1
\]

It’s understood that the fraction is \(d(x) + 1\) if \(t = 1\).

(b) For the graph \((S, \uparrow)\),

\[
\Gamma(x, t) = \frac{t^{d(x)+1} - (1-t)^{d(x)+1}}{(1-t)^{d(x)+1}(2t-1)}, \quad x \in S, \ |t| < 1
\]

(c) For the graph \((S, \prec)\),

\[
\Gamma(x, t) = (1 + t)^{d(x)}, \quad x \in S, \ t \in \mathbb{R}
\]
Exercise 13.3  Suppose that random variable $X$ takes values in $S$ and that $n \in \mathbb{N}$.

(a) For the graph $(S, \uparrow)$,
\[ \sum_{d(x) \geq n} P(x \uparrow X) = P[d(X) \geq n + 1] \]

(b) For the graph $(S, \uparrow)$,
\[ \sum_{x \in S} \sum_{k=0}^{d(x)} \binom{n}{k} P(x \uparrow X) = E \left[ \sum_{k=0}^{d(X)} \binom{n+1}{k} \right] \]

(c) For the graph $(S, \prec)$,
\[ \sum_{x \in S} \binom{d(x)}{n} P(X > x) = E \left[ \binom{d(X)}{n+1} \right] \]

Chapter 14

Exercise 14.1  Recall that $\tau_k$ denotes the left walk function of order $k \in \mathbb{N}$ for the graph $(\mathbb{N}_+, \preceq)$.

(a) For the graph $(\mathbb{N}_+, \prec)$, the left walk function $\hat{\tau}_k$ of order $k$ is given by
\[ \hat{\tau}_k(x) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} \tau_j(x), \quad x \in \mathbb{N}_+ \]

This follows from Proposition 14.2 and Theorem 6.1

(b) For the graph $(\mathbb{N}_+, \uparrow)$, The left walk function $\gamma_k$ of order $k$ is given by
\[ \gamma_k(x) = \sum \binom{k}{y} : y \in \mathbb{N}_+, y \preceq x, \sum_{i \in I} n_i(y) = k \]

This follows by a simple combinatorial argument

(c) For the graph $(\mathbb{N}_+, \uparrow)$, the left walk function $\hat{\gamma}_k$ of order $k$ is given by
\[ \hat{\gamma}_k(x) = \sum_{j=0}^{k} \binom{k}{j} \gamma_j(x), \quad x \in \mathbb{N}_+ \]

This follows from (b) and Theorem 6.1.

Chapter 15

Exercise 15.1  Let $m = 2$ corresponding to the path on $(0, 1, 2)$. Define $f$ and $g$ by
\[ f(0) = \frac{1}{4}, \quad f(1) = \frac{1}{2}, \quad f(2) = \frac{1}{4} \]
\[ g(0) = \frac{1}{8}, \quad g(1) = \frac{1}{2}, \quad g(2) = \frac{3}{8} \]

Then $f$ and $g$ are distinct probability density functions on $S$ that generate the same right probability function $F$, given by $F(x) = \frac{1}{2}$ for $x \in \{0, 1, 2\}$. 
Exercise 15.2  Let $m = 2$ corresponding to the path $(0, 1, 2)$.

(a) The left generating function $\Gamma$ is given by

\[
\Gamma(0, t) = \Gamma(2, t) = \frac{1 + t}{1 - 2t^2}
\]
\[
\Gamma(1, t) = \frac{1 + 2t}{1 - 2t^2}
\]

(b) Let $k \in \mathbb{N}_+$. The left walk function $\gamma_{2k}$ of even order $2k$ is given by

\[
\gamma_{2k}(x) = 2^k, \quad x \in S
\]

The left walk function $\gamma_{2k+1}$ of odd order $2k + 1$ is given by

\[
\gamma_{2k+1}(0) = \gamma_{2k+1}(2) = 2^k
\]
\[
\gamma_{2k+1}(1) = 2^{k+1}
\]

(c) The constant rate distribution has rate $\alpha_2 = \frac{1}{\sqrt{2}} \approx 0.707$. The density function $f$ of this distribution is

\[
f(0) = f(2) = \frac{1}{2 + \sqrt{2}} \approx 0.293
\]
\[
f(1) = \frac{\sqrt{2}}{2 + \sqrt{2}} \approx 0.414
\]

The mean is $\mu = 1$ and the variance is $\sigma^2 = 2 - \sqrt{2} \approx 0.586$.

(d) Let $Y = (Y_1, Y_2, \ldots)$ be the random walk on $(S, \rightarrow)$ corresponding to the constant rate distribution. For $k \in \mathbb{N}_+$, the density function $f_{2k}$ of $Y_{2k}$ is given by

\[
f_{2k}(0) = f_{2k}(2) = \frac{1}{\sqrt{2}} f(0) = \frac{1}{2 + 2\sqrt{2}} \approx 0.207
\]
\[
f_{2k}(1) = \sqrt{2} f(1) = \frac{2}{2 + \sqrt{2}} \approx 0.586
\]

The density function $f_{2k+1}$ of $Y_{2k+1}$ is $f_{2k+1} = f$. So the distribution of $Y_n$ depends on $n \in \mathbb{N}_+$ only through parity (odd or even).

Exercise 15.3  When $m = 3$ corresponding to the path $(0, 1, 2, 3)$.

(a) The left walk functions have an interesting representation in terms of the Fibonacci numbers. For $n \in \mathbb{N}$, let $c_n$ denote the $n$th Fibonacci number, so that

\[
c = (0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots)
\]

For $n \in \mathbb{N}$,

\[
\gamma_n(0) = \gamma_n(3) = c_{n+1}
\]
\[
\gamma_n(1) = \gamma_n(2) = c_{n+2}
\]

(b) The left generating function is given by

\[
\Gamma(0, t) = \Gamma(3, t) = \frac{1}{1 - t - t^2}
\]
\[
\Gamma(1, t) = \Gamma(2, t) = \frac{1 + t}{1 - t - t^2}
\]
(c) The constant rate distribution has rate $\alpha = \frac{\sqrt{5} - 1}{2} \approx 0.618$. The PDF $f$ is given by

\[
    f(0) = f(3) = \frac{3 - \sqrt{5}}{4} \approx 0.191 \\
    f(1) = f(2) = \frac{\sqrt{5} - 1}{4} \approx 0.309
\]

The mean is $\mu = \frac{3}{2}$ and the variance is $\sigma^2 = \frac{13}{4} - \sqrt{5} = 1.014$.

(d) The Fibonacci numbers, and hence the cumulative function can be written in terms of the constant rate. For $n \in \mathbb{N}$,

\[c_n = \frac{\alpha_3^{-n} + (-1)^{n+1} \alpha_3^n}{2\alpha_3 + 1}\]

This is essentially Binet’s formula.

(e) The gamma PDF of order $n \in \mathbb{N}_+$ for $(S, \rightarrow)$ is given by

\[
    f_n(0) = f_n(3) = \frac{\alpha_3 + (-1)^{n+1} \alpha_3^2 + 1}{4\alpha_3 + 2} \\
    f_n(1) = f_n(2) = \frac{1 + \alpha_3 + (-1)^n \alpha_3^2 + 1}{4\alpha_3 + 2}
\]

Note that

\[
    f_n(0) = f_n(3) \rightarrow \frac{\alpha_3}{4\alpha_3 + 2} = \frac{5 - \sqrt{5}}{20} \approx 0.138 \text{ as } n \rightarrow \infty \\
    f_n(1) = f_n(2) \rightarrow \frac{1 + \alpha_3}{4\alpha_3 + 2} = \frac{5 + \sqrt{5}}{20} \approx 0.362 \text{ as } n \rightarrow \infty
\]

**Exercise 15.4** Let $m = 4$, corresponding to the path $(0, 1, 2, 3, 4)$.

(a) The left walk function can be given in closed form. For $k \in \mathbb{N}_+$,

\[
    \gamma_{2k}(0) = \gamma_{2k}(4) = 2 \cdot 3^{k-1}, \quad \gamma_{2k+1}(0) = \gamma_{2k+1}(4) = 3^k \\
    l_{2k}(1) = l_{2k}(3) = 3^k, \quad l_{2k+1}(1) = l_{2k+1}(3) = 2 \cdot 3^k \\
    \gamma_{2k}(2) = 4 \cdot 3^{k-1}, \quad \gamma_{2k+1}(2) = 2 \cdot 3^k
\]

(b) The left generating function is

\[
    \Gamma(0, t) = \Gamma(4, t) = \frac{1 + t - t^2}{1 - 3t^2} \\
    \Gamma(1, t) = \Gamma(3, t) = \frac{1 + 2t}{1 - 3t^2} \\
    \Gamma(2, t) = \frac{(1 + t)^2}{1 - 3t^2}
\]

(c) The constant rate distribution has rate $\alpha_4 = \frac{1}{\sqrt{3}} \approx 0.577$. The PDF $f$ of this distribution is

\[
    f(0) = f(4) = \frac{1}{4 + 2\sqrt{3}} \approx 0.134 \\
    f(1) = f(3) = \frac{3}{6 + 4\sqrt{3}} \approx 0.232 \\
    f(2) = \frac{1}{2 + \sqrt{3}} \approx 0.268
\]

The mean is $\mu = 2$ and the variance is $\sigma^2 = 5 - 2\sqrt{3} \approx 1.536$. 

(d) Let \( k \in \mathbb{N}_+ \). The gamma distribution of order \( 2k \) is given by

\[
\begin{align*}
f_{2k}(0) &= f_{2k}(4) = \frac{1}{6 + 4\sqrt{3}} \approx 0.0773 \\
f_{2k}(1) &= f_{2k}(3) = \frac{1}{2 + \sqrt{3}} \approx 0.268 \\
f_{2k}(2) &= \frac{3}{3 + 2\sqrt{3}} \approx 0.309
\end{align*}
\]

The gamma distribution of order \( 2k + 1 \) is given by

\[
\begin{align*}
f_{2k+1}(0) &= f_{2k+1}(4) = \frac{1}{6 + 3\sqrt{3}} \approx 0.0893 \\
f_{2k+1}(1) &= f_{2k+1}(3) = \frac{3}{6 + 4\sqrt{3}} \approx 0.232 \\
f_{2k+1}(2) &= \frac{4}{6 + 3\sqrt{3}} \approx 0.357
\end{align*}
\]

So there are three distinct gamma distribution, corresponding to \( n = 1, \) odd \( n \geq 3 \) and even \( n \).

Chapter 16

Exercise 16.1 \( x = \{2, 5, 6, 8, 12, 13, 25\}, \ y = \{1, 4, 7\}, \ z = \{1, 3, 4, 6\} \)

(a) \( x \circ y = \{2, 8, 25\}, \ x \circ z = \{2, 6, 8, 13\} \)

(b) \( x \circ (y \cup z) = (x \circ y) \cup (x \circ z) = \{2, 6, 8, 13, 25\} \)

(c) \( x \circ (y \cap z) = (x \circ y) \cap (x \circ z) = \{2, 8\} \)

Exercise 16.2 \( x = \{1, 3, 5, \ldots\}, \ y = \{2, 4, 6, \ldots\} \)

(a) \( x \circ x = \{1, 5, 9, \ldots\} \)

(b) \( x \circ y = \{3, 7, 11, \ldots\} \)

(c) \( y \circ x = \{2, 6, 10, \ldots\} \)

(d) \( y \circ y = \{4, 8, 12, \ldots\} \)

Exercise 16.3 \( x = \{2, 3, 10\}, \ y = \{4, 7\} \)

(a) \( xy = \{2, 3, 6, 8, 10\} \)

(b) \( yx = \{2, 3, 4, 7, 12\} \)

(c) \( x^2 = \{2, 3, 4, 5, 10, 13\} \)

(d) \( y^2 = \{4, 5, 7, 9\} \)

Exercise 16.4  Factorings of two and three element sets.

(a) Suppose that \( i, j \in \mathbb{N}_+ \) with \( i < j \). Then \( \{i, j\} = \{j\}\{i\} = \{i\}\{j - 1\} \).

(b) Suppose that \( i, j, k \in \mathbb{N}_+ \) with \( i < j < k \). Then

\[
\begin{align*}
\{i, j, k\} &= \{k\}\{j\}\{i\} = \{k\}\{i\}\{j - 1\} = \{j\}\{i\}\{k - 2\} \\
&= \{j\}\{k - 1\}\{i\} = \{i\}\{k - 1\}\{j - 1\} = \{i\}\{j - 1\}\{k - 2\}
\end{align*}
\]
**Exercise 16.5** Let $x = \{2, 3, 5, 12, 17\}$. Factor $x$ into singletons corresponding to the given permutations:

(a) For the permutation $(3, 5, 17, 2, 12)$, $x = \{3\} \{4\} \{15\} \{2\} \{9\}$.

(b) For the permutation $(2, 5, 3, 17, 12)$, $x = \{2\} \{4\} \{2\} \{14\} \{9\}$.

**Exercise 16.6** The set of minimal elements of $(S_k, \cdot)$ for $k \in \{1, 2, 3\}$

(a) $M_1 = \{\{2\}, \{1, 3\}\}$

(b) $M_2 = \{\{3\}, \{1, 4\}, \{2, 4\}, \{1, 2, 5\}\}$

(c) $M_3 = \{\{4\}, \{1, 5\}, \{2, 5\}, \{3, 5\}, \{1, 2, 6\}, \{1, 3, 6\}, \{2, 3, 6\}, \{1, 2, 3, 7\}\}$

**Exercise 16.7**

$$G(\{2, 4, \ldots\}) = \frac{pr^2}{1 - r^2 + pr^2}$$

**Exercise 16.8**

$$G(\{1, 3, \ldots\}) = \frac{rp}{1 - r^2 + pr^2}$$

**Chapter 17**

**Exercise 17.1** See Figure B.1.

![Figure B.1: Lexicographic graphs: (a) $k_0 = k_1 = k_2 = 2, \ldots$ (b) $k_0 = 1, k_1 = 2, k_2 = 3, \ldots$](image)

**Exercise 17.2** The walk function $\gamma_m$ of order $m \in \mathbb{N}_+$ for $(S, \uparrow)$ is given by

$$\delta_m(x) = \sum_{j=0}^{m} \binom{m}{j} a_{n-j} a_{n-j+1} \cdots a_{n-1}, \quad x \in S_n, n \in \mathbb{N}$$

**Exercise 17.4** Let $m \in \mathbb{N}_+$. Then $Y_m$ has right probability function $F_m$ for $(S, \uparrow)$ given by

$$F_m(x) = \mathbb{P}(Y_m \in S_{n+1}) = \mathbb{P}(N_m = n + 1) = \frac{1}{\alpha^{n-m+2} k_0 \cdots k_{n-m+1}}, \quad x \in S_n, n \in \{m - 1, m, m + 1, \ldots\}$$

Hence $Y_m$ has right rate function $r_m = f_m/F_m$ for $(S, \uparrow)$ given by $r_m(x) = 0$ for $x \in S_n$ with $n \in \{0, \ldots, m-2\}$ and

$$r_m(x) = \alpha \frac{k_{n-m+1}}{k_n}, \quad x \in S_n, n \in \{m - 1, m, m + 1, \ldots\}$$
Exercise 17.6 The probability generating function is given by

\[ E(t^N) = \frac{te^t - e^t + 1}{t^2}, \quad t \in \mathbb{R} \]

The mean is \(E(N) = e - 1 \approx 1.718\) and the variance is \(\text{var}(N) = 4e - e^2 \approx 3.481\).

Exercise 17.7 Memoryless probability density functions in the lexicographic semigroup with \(k = 4\) and \(q = \frac{1}{2}\). So each has upper probability function \(F\) given by \(F(i^n) = \left(\frac{1}{2}\right)^n\) for \(i^n \in S\). Only the distribution in part (b) is exponential.

(a) \(f(e) = 0\). The probability density function \(f\) is given by

\[ f(i^n) = \frac{1}{5} \left[ \left(\frac{1}{2}\right)^n - \left(-\frac{1}{3}\right)^n \right], \quad i^n \in S \]

(b) \(f(e) = \frac{1}{5}\). The probability density function \(f\) is given by

\[ f(i^n) = \frac{1}{5} \left(\frac{1}{2}\right)^n, \quad i^n \in S \]

(c) \(f(e) = \frac{2}{5}\). The probability density function \(f\) is given by

\[ f(i^n) = \frac{1}{5} \left[ \left(\frac{1}{2}\right)^n + \left(-\frac{1}{3}\right)^n \right], \quad i^n \in S \]

(d) \(f(e) = \frac{1}{2}\). The probability density function \(f\) of is given by

\[ f(i^n) = \frac{1}{5} \left(\frac{1}{2}\right)^n + \frac{3}{10} \left(-\frac{1}{3}\right)^n, \quad i^n \in S \]

Note that \(f(i) = 0\) for \(i \in I\).
Bibliography


[29] M Pendergrass, private communication