NETWORKS GOES TO SCHOOL

#2

for secondary schools

NET WORKS

NET
WORKS

NETWORKS is a 10-year research programme hosted by the University of Amsterdam (UvA), Eindhoven University of Technology (TU/e), Leiden University (UL), and the Center for Mathematics and Computer Science (CWI) in Amsterdam.

The work in this book is supported by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research (NWO) through Gravitation-grant NETWORKS-024.002.003.



CONTENTS

INTRODUCTION

Transportation, traffic, communication and energy networks form the backbone of our modern society. To deal with the uncertainty, variation, unpredictability, size and complexity inherent in these networks, we need to develop radically new ways of thinking. The ultimate goal is to build self-organizing and intelligent networks. The NWO-funded Gravitation programme NETWORKS started in the Summer of 2014 and covers a broad range of topics dealing with stochastic and algorithmic aspects of networks.

In March 2019 the second "NETWORKS goes to school" event was organized. The aim of the event is to provide secondary education students and teachers a first mathematical introduction on network sciences. This book collects the material realised for the "NETWORKS goes to school" event.

In Chapter 1 all the necessary background material is presented; this can be used as a starting point before reading Chapters 2. In Section 2.1 we present some methods and results from queueing theory. We show how we can use mathematics to study a queueing system, some techniques that are broadly used and finally we present some important, almost 'classical', results.
Section 2.2 focuses on mathematical genetics. We show how probability theory can be applied in this field and attention is paid to some basic methods and some important results.
Chapter 2 has been written with the help of Mark van der Boor (Eindhoven University of Technology) and Margriet Oomen (Leiden University).

For more information and the book of the first "NETWORKS goes to school", please visit www.networkpages.nl

On behalf of the NETWORKS programme,

the organising committee of "NETWORKS goes to school"
Nicos Starreveld (University of Amsterdam)
Marta Maggioni (Leiden University)

CHAPTER 1

Mathematical background

In this chapter the necessary background knowledge is provided. In Section 1.2 a random variable is defined, and some basic concepts from probability theory together with some examples are presented. In Section 1.3 the hierarchical lattice, a specific type of network, is defined.

1.1. Basic notation

We start by introducing some notation we will use in the sequel:

(1) $\mathbb{N}$ for the set of natural numbers, that is $\mathbb{N} = \{1, 2, 3, \cdots\}$;
(2) $\mathbb{N}_0$ for the set of natural numbers included zero, that is $\mathbb{N}_0 = \{0, 1, 2, 3, \cdots\}$;
(3) $\mathbb{Z}$ for the set of integer numbers, that is $\mathbb{Z} = \{\cdots, -3, -2, -1, 0, 1, 2, 3, \cdots\}$.

1.2. Probability theory

Probability theory is the area of mathematics that studies random phenomena. For example if the experiment is tossing a coin, then there are two possible outcomes, either *heads* or *tails*. Each outcome occurs with probability 0.5. In order to study such a random experiment we use random variables.

Random variable

A **random variable** X is a variable whose possible values are outcomes of a random experiment.

We define a random variable by giving the *state space*, i.e. the set of all possible values the variable can take, and the *probability function*, which yields the corresponding probability that a given outcome will occur. For the coin toss for example we can define a random variable by assigning to the outcome *heads* the value 1 and to the outcome *tails* the value 0. In this case we have

$$X(\textit{heads}) = 1 \quad \text{and} \quad X(\textit{tails}) = 0.$$

The probability function for this random variable is given by

$$\mathbb{P}(X = 1) = \mathbb{P}(\textit{heads}) = 0.5,$$

and

$$\mathbb{P}(X = 0) = \mathbb{P}(\textit{tails}) = 0.5,$$

where for a possible set of outcomes A, $\mathbb{P}(A)$ denotes the probability that A occurs. A random variable can be *discrete* or *continuous*.

Discrete random variables

A random variable X is called discrete when it can take countable many values, for simplicity we can just say that its values are the integer numbers, that is $X \in \mathbb{Z}$.

Continuous random variables

A random variable X is called continuous when it can take continuously many values, for simplicity we can just say that its values are the real numbers that is $X \in \mathbb{R}$.

For a discrete random variable, we can write down the probability that it equals a specific value. For a continuous random variable, this is not possible, as there is a continuum of possible values. We can however specify the probability that a continuous random variable falls in a range of values by using the **density function**. The probability that a continuous random variable X assumes a value in $[a, b]$ is given by the integral of the density function over that interval. This corresponds to the area delimited by the graph of the density function, the x-axis and the vertical lines $y = a$ and $y = b$.

1.2.1. Bernoulli random variable

Bernoulli random variable

A **Bernoulli random variable** describes the outcome of any single random experiment that asks a yes-no question, like tossing a coin.

It takes the value 1 with probability p and the value 0 with probability $1 - p$. Consider for example a coin where one side is heavier, then this is a biased coin where one side is favoured. We will use $B(p)$ to denote a Bernoulli random variable with probability p.

1.2.2. Binomial random variable

Binomial random variable

A **binomial random variable** describes the number of successes in a series of experiments.

We make the following assumptions:

- the number n of observations is fixed;
- each observation is independent of the other observations;
- each observation represents one of two outcomes: success or failure;
- the probability p of success is exactly the same for each trial.

Under these assumptions, we can describe each binomial distribution by using the parameters n and p, we will denote such a binomial random variable by $B(n, p)$.
Let X be a random variable that follows the binomial distribution with parameters n and p. Then X has state space $\{0, 1, \ldots, n\}$, and the probability that X is equal to k is given by

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

where

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

is the **binomial coefficient**. The symbol $\binom{n}{k}$ is read as 'n choose k', as this is the number of ways to choose k different elements from a total of n elements, where the order of elements does not matter. The factorial of n is denoted by $n!$ and it is equal to the product $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$.

EXAMPLE 1.2.1. Suppose that we have a total of 5 colours, and we wish to know how many combinations there are of 3 different colours, where the order of the colours does not matter. Then $n = 5$ and $k = 3$, and

$$\binom{5}{3} = \frac{5!}{3!2!} = 10.$$

We could also reason in a different way. For the first choice we have a total of 5 possible colours, for the second choice we have 4 possible colours and for the third choice we have 3 possible colours. The total of combinations of three colours is then $5 \cdot 4 \cdot 3 = 5!/2!$. However, the order of colours did not matter so we still have to divide by the number of ways in which we can order 3 colours, which is $3 \cdot 2 \cdot 1 = 3!$.

EXAMPLE 1.2.2. Consider a coin toss, where possible outcomes are heads or tails. Suppose that we have a fair coin, i.e., the probability for heads is the same as it is for tails. If we toss the coin 10 times, then the number of coin tosses that came heads from those ten tosses has a binomial distribution with parameters $n = 10$ and $p = \frac{1}{2}$. The probability of getting exactly four heads is equal to

$$\mathbb{P}(X = 4) = \binom{10}{4} \frac{1}{2}^4 \left(1 - \frac{1}{2}\right)^{10-4} = \frac{105}{512} \approx 0.205.$$

1.2.3. Geometric random variable

Geometric random variable

A **geometric random variable** describes the number of failures in a sequence of simple experiments until the first success.

We make the following assumptions:

- each observation is independent of the other observations;
- each observation represents one of two outcomes: success or failure;
- the probability p of success is exactly the same for each trial.

Under these assumptions, we can describe each geometric distribution by using the parameter p, we will denote a geometric random variable by $G(p)$. Let X be a random variable that follows the geometric distribution with parameter p. Then X has state space $\{0, 1, 2, \ldots\}$, and the probability that X is equal to k is given by

$$\mathbb{P}(X = k) = (1 - p)^k p.$$

When the event random variable X is equal to k then we know that k failures have occurred before the first success. The probability of failure is equal to $1 - p$ and by the assumptions above the experiments we perform are independent of each other.

EXAMPLE 1.2.3. Consider a coin toss, where possible outcomes are heads or tails. Suppose that we have an unfair coin, i.e., the probability for heads is $\frac{2}{3}$ and the probability for tails is $\frac{1}{3}$. Then the probability to get 5 times tails before the first heads is equal to

$$\mathbb{P}(X = 5) = \left(\frac{1}{3}\right)^5 \frac{2}{3} \approx 0.0027.$$

1.2.4. Exponential random variable

Exponential random variable

The **exponential random variable** is a continuous random variable and is usually encountered when we study the occurrences of consecutive independent events.

An exponential random variable is characterised by a parameter λ, called the intensity. The larger this parameter is the higher the frequency of the arriving events. A random variable having the exponential distribution with parameter λ, denoted by $E(\lambda)$, has the following probability distribution function

$$\mathbb{P}(E(\lambda) \leq x) = 1 - e^{-\lambda x}, \qquad x \geq 0, \tag{1.2.1}$$

and a probability density function given by

$$f_\lambda(x) = \lambda e^{-\lambda x}. \tag{1.2.2}$$

As seen before for the geometric random variable, the exponential random variable has the memoryless property, i.e. that means that

$$\mathbb{P}(E(\lambda) > x + y | E(\lambda) > y) = \mathbb{P}(E(\lambda) > x), \qquad x, y \geq 0. \tag{1.2.3}$$

This memoryless property is quite remarkable, so let's look at it from a practical side. Suppose the time until the bus arrives is exponentially distributed. If that would be the case, then if the bus didn't arrive for an hour, then it would still take the same amount of time until the bus arrives. But in reality we expect that if the bus didn't arrive for an hour, then it will probably arrive soon.

1.2.5. Poisson process

Finally, we introduce the Poisson process. This is a process where events happen once every while. The time between events is exponentially distributed. Since the exponential distribution is memoryless, the Poisson process has a very remarkable property. If no event happened for a while, it doesn't imply that some event will occur soon. As an example, consider the time until you hit a specific number on a roulette wheel. If that specific number didn't show up for a while, that doesn't make it more likely for the number to show up sooner than normal. In other words: the history of the process has no influence on the future.

1.3. The hierarchical lattice

The hierarchical lattice is a type of a network with a very specific structure. The hierarchical lattice is characterised by a parameter N and resembles a tree whose branches can keep branching infinitely often.

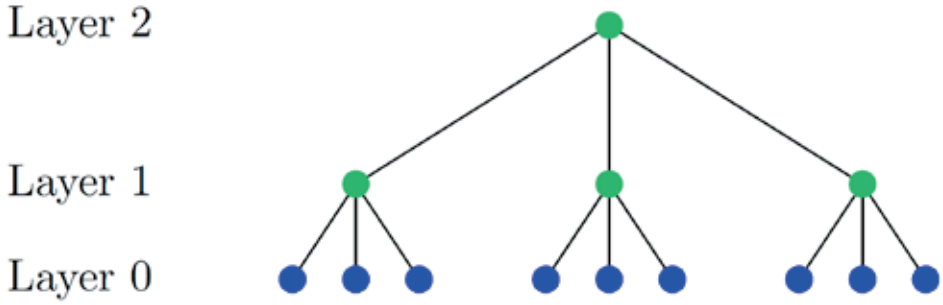
Tree

A tree is defined as a graph which has no cycles. In such a graph any two points are connected by exactly one path. In the figures below you can find some examples of trees. The points in a tree that are connected with only one other point are called the leaves of the tree.

Let's see in steps how the hierarchical lattice is constructed. As said before the hierarchical lattice consists of infinitely many layers. The first step is to see what happens at the the bottom layer, think of the bottom layer as the leaves of the tree. We need to define how many leaves, or nodes, we will place at the bottom layer, say d. For simplicity we illustrate the steps for the case we have $d = 3$, that is 3 leaves at the bottom layer. We start thus by placing 3 points in layer 0

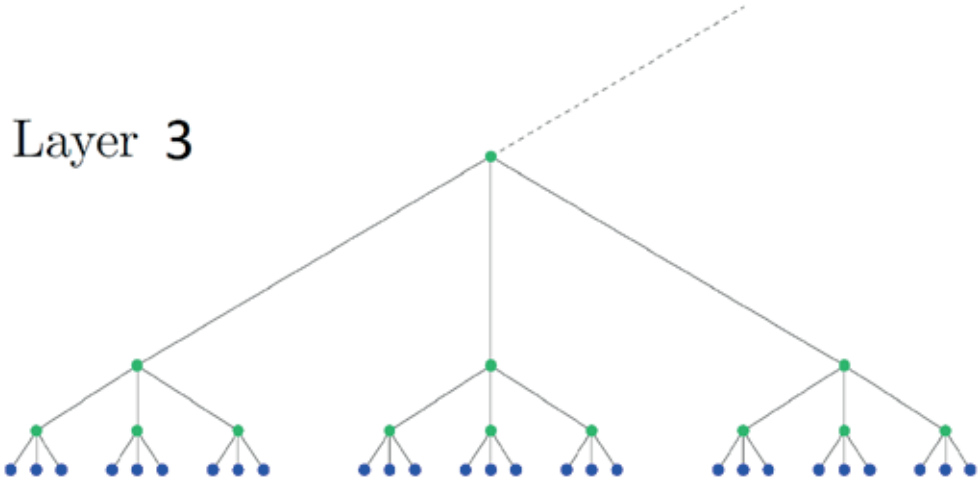
Afterwards we connect these points with a single ancestor (second generation). This way we obtain the first point in layer 1.
In the next step we create d, in our case $d = 3$, copies of this graph, this yields

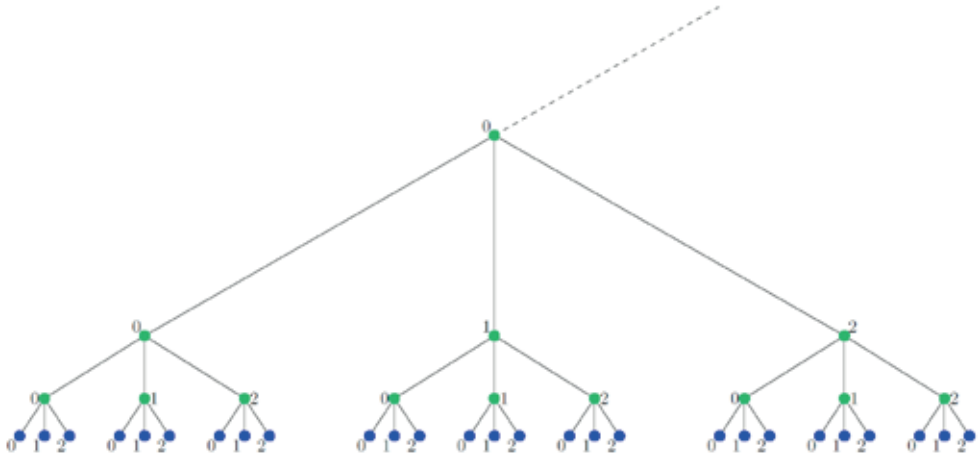
Connect the second generation point, those are the points in layer 1, with a third generation point, this yields the first point in layer 2.



Follow the same steps to create the higher order generations.

We observe that building the network according to this algorithmic procedure leads to infinitely many points in layer zero. The hierarchical lattice consists only of the points in layer zero, which are the blue points in the figures; the ancestors, which are the green points, are the branching points that allow us to build the graph. A hierarchical lattice with parameter d will be denoted by Ω_d, in our case Ω_3. We observe that the way to construct the graph is rather simple. But we would also like to have an efficient way to characterise all the points belonging to Ω_3. We do this by giving a label to each point based on its position in the lattice. We look at layer i, for $i = 0, 1, 2, \ldots$ and starting from the left we assign to consecutive points the numbers $0, 1, 2$. Hence we obtain the following numbering


Layer 3

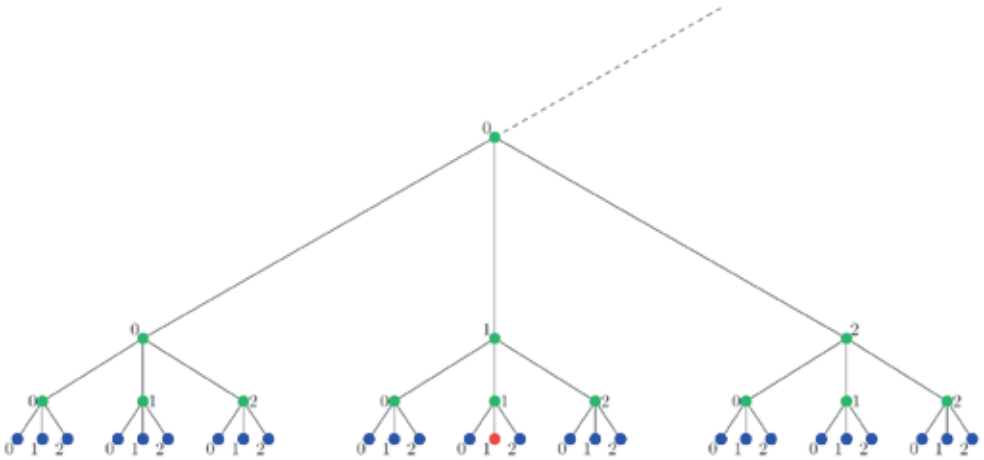


Label of a point in the hierarchical lattice

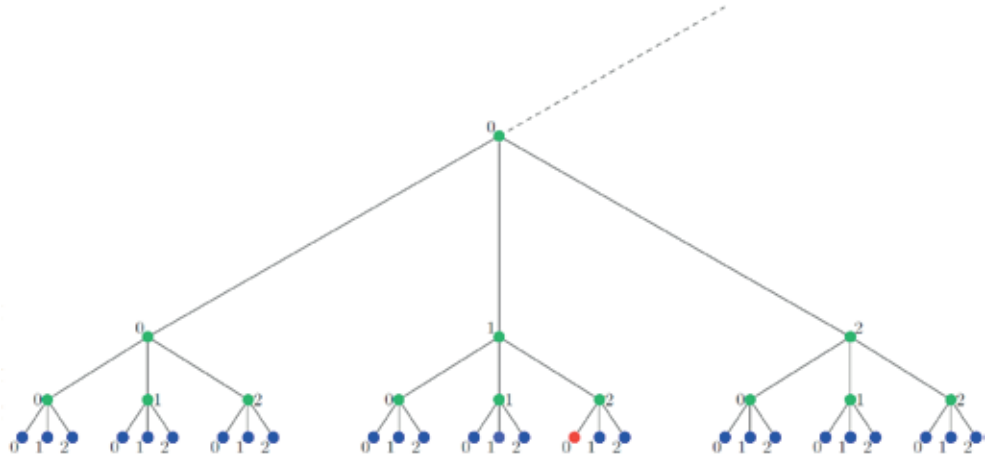
Every point in the lattice is then identified with an infinitely long sequence where each element can take one of the values $0, 1, 2$. The j-th element in the sequence gives the position in the corresponding block of length 3 in layer j.

EXAMPLE 1.3.1. Let's identify the following points:

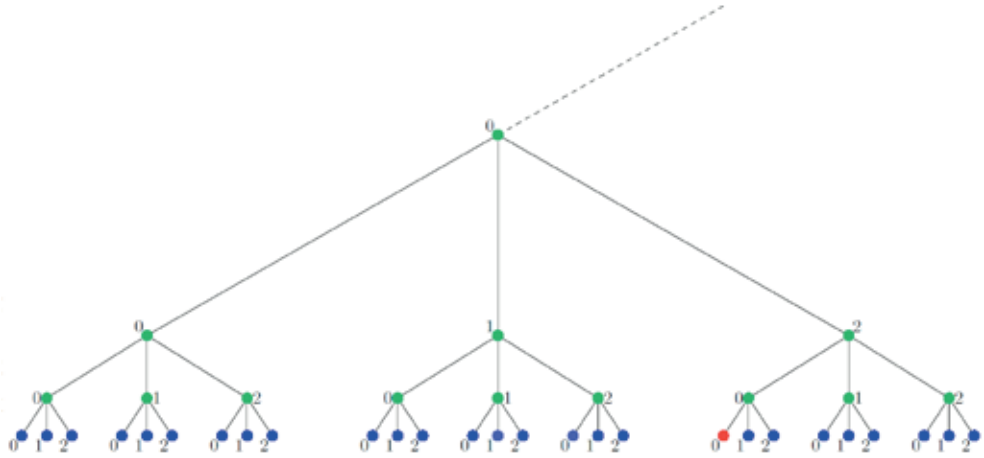
- The sequence $(1, 1, 1, 0, 0, 0, \ldots)$ corresponds to the point



- The sequence $(0, 2, 1, 0, 0, 0, \ldots)$ corresponds to the point



- The sequence $(0, 0, 2, 0, 0, 0, \ldots)$ corresponds to the point



The formal characterisation of the hierarchical lattice is given below

$$\Omega_3 = \left\{ \xi = (\xi_n)_{n\in\mathbb{N}_0} : \xi_n \in \{0,1,2\}, \sum_{n\in\mathbb{N}_0} \xi_n < \infty \right\}. \tag{1.3.1}$$

This set contains exactly all sequences with only a finite amount of positive entries, which is due to the fact that each point has a label of finite length. These definitions may seem rather intimidating so lets break them down into pieces. Lets see how an element $\xi = (\xi_n)_{n\in\mathbb{N}_0}$ looks like. The notation $(\xi_n)_{n\in\mathbb{N}_0}$ refers to a sequence of infinite length, that is

$$(\xi_n)_{n\in\mathbb{N}_0} = (\xi_0, \xi_1, \xi_2, \dots,).$$

We need to define which are the possible values of the elements of this sequence, i.e. of ξ_n. Each one of these elements can be equal to one of the numbers $0, 1, 2$.

Examples of sequences

The elements

- $\xi = (0,0,0,0,\dots,0,\dots)$
- $\eta = (1,0,0,0,\dots,0,\dots)$
- $\zeta = (1,4,2,2,3,1,0,4,3,1,0,0,\dots,0,\dots)$
- $\psi = (4,4,4,4,0,0\dots,0,\dots)$

all belong to Ω_3.

What we observe in all these element is that after some position in the sequence all the elements that come afterwards are equal to zero. This is the third condition we impose on elements from Ω_3, that after some position in the sequence all the elements that follow are equal to zero. This is written in abstract mathematical language as

$$\sum_{n\in\mathbb{N}_0} \xi_n < \infty.$$

Try to understand the interpretation of this condition. The next step is to define a distance between points on the hierarchical lattice Ω_3, you would like to know when are point considered to be far from each other and when close to each other. The distance function between the points ξ and η is given by

$$d(\xi,\eta) = \min\{m \in \mathbb{N}_0 : \xi_n = \eta_n, \forall n \geq m\}. \tag{1.3.2}$$

EXAMPLE 1.3.2. In this example we will compute the distance of some elements from Ω_3.

- $\xi = (0,0,0,0,\dots)$, i.e. the first point on the left, and $\eta = (1,0,0,0,\dots)$. Then $d(\xi,\eta) = 1$.
- $\xi = (0,0,0,0,\dots)$ and $\eta = (0,1,0,0,\dots)$. Then $d(\xi,\eta) = 2$.
- $\xi = (2,0,0,0,\dots)$ and $\eta = (0,1,0,0,\dots)$. Then $d(\xi,\eta) = 2$.
- $\xi = (0,0,0,0,\dots)$ and $\eta = (0,0,1,0,\dots)$. Then $d(\xi,\eta) = 3$.
- $\xi = (0,0,1,0,\dots)$ and $\eta = (1,0,1,0,\dots)$. Then $d(\xi,\eta) = 1$.

CHAPTER 2

Queueing theory and networks in genetics

In this chapter some results from queueing theory and mathematical genetics are presented. Section 2.1 is about queueing theory, a branch of probability theory that studies cases and systems where there is demand for some scarce resource. Section 2.2 concerns mathematical genetics and the mechanisms that govern the evolution of genes.

2.1. Queueing theory – Waiting in an efficient way

Queueing theory

Queueing theory refers to the branch of probability theory that studies cases and systems where there is demand for some scarce resource.

The main characteristics of such systems relate to clients who arrive to, possibly multiple, servers and receive some form of service. After being served a client may depart from the system or proceed to a next service station. Neither the customers nor the servers are necessarily actual individuals; the clients may be objects, phone calls, orders for some product while the servers may be computer programs or machines. Some typical everyday examples of queueing systems can be found in supermarkets, industrial production systems and hospitals. In a supermarket customers arrive to the counters, they may have to wait in the queue until their turn comes, they are served and then leave the supermarket. In an industrial production system, like a factory producing cars, the products have to undergo multiple stages until they are assembled and the servers may be either machines or individuals. Finally, patients arriving to a hospital often need access to resources like doctors, beds, medicine and equipment. The time a patient spends in the hospital using medical resources and possibly occupying a bed, constitutes his service time. A new patient can go into treatment only when the hospital has the necessary resources available, for example only if there are free beds.

Performance measures

The performance of queueing systems is expressed using *performance measures*. The most commonly used performance measures are the *queue length* and the *waiting time*.

2.1.1. Mathematical model of a queue

A mathematical model

When we speak of a mathematical model we mean a description of a system or some real life situation, in this case a queue, using mathematical concepts.

Setting up a suitable mathematical model for an application is an important part of doing mathematics. A mathematical model describes real life situations in such a way that it enables you to do computations. In this section we describe the basic principles used in order

to construct a mathematical model for a queueing system. At an abstract level a model of a queueing system has to take into account the following quantities:

- the time between consequent arriving customers, called the *interarrival time*;
- the time the server needs to serve a customer, called the *service time* of that customer;
- the number of servers.

And that is not all! More difficult queueing systems might also have limited waiting rooms, the servers might be connected in a network and customers may leave the system if they are in there for too long.

In queueing systems congestion, or long queues, typically appear when many customers join the system in a short time, or if the server takes long to serve the customers: this is all caused by the randomness in the service requirements and the arrival process of customers. Hence queueing models are typically of a probabilistic nature, and the interarrival and service times are considered to have some known probability distribution. To understand the behavior of a system and to optimise its performance, it is important to take the randomness into account, starting with determining appropriate probability distributions for the interarrival and service times.

In what follows we describe the simplest mathematical model of a queueing system, called the M/M/1 model. In the M/M/1 model

- the time between consecutive arriving customers has an exponential distribution with parameter λ (or equivalently, customers arrive according to a Poisson process with rate λ);
- the time the server needs to serve a customer has an exponential distribution with parameter μ;
- there is one server present in the system.

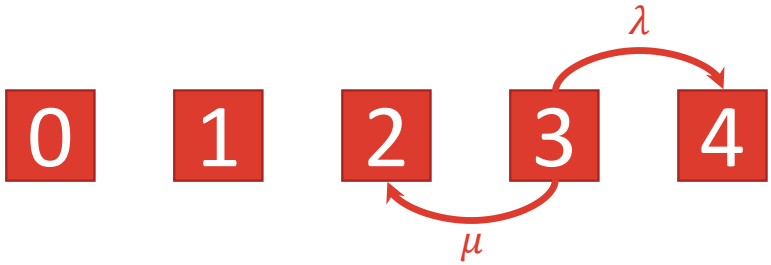
In the figure below we show this using an illustration.

Such illustrations are typical when representing queueing systems, we use an incoming arrow for the stream of arriving customers, a circle for each server and a (blue) line for each customer in the queue. In the M/M/1 model we suppose that the system has infinite space so infinitely many customers can wait in the queue. Denote by Q a random variable describing the number of customers in the system, that is the number of customers waiting in the queue plus the customer, if the system is not empty, that is being served. Then Q has as state space the set $\{0, 1, 2, \ldots\}$. We find the probability distribution of Q using an argument that is called *flow conservation argument*. Let's see how it works. We start by drawing a diagram to denote the number of jobs in the system, such a diagram is depicted in the figure below.

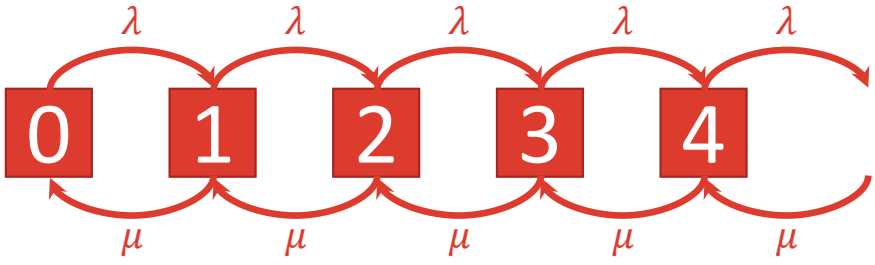
Each square represents a possible state of the system, for example the square with 0 represents the state that there are no customers in the system.

2.1.2. Flow conservation argument

Suppose that $Q = i$, that is there are i customers in the system. Then two things can occur, the customer who is being served departs from the system before a new customer arrives, or a new customer arrives before the customer who is being served departs the system. The first event corresponds to the transition $\{Q = i\} \to \{Q = i - 1\}$ since a customer departs. The service time has an exponential distribution with parameter μ, hence the transition $\{Q = i\} \to \{Q = i - 1\}$ occurs with rate μ. On the other side the second event corresponds to the transition $\{Q = i\} \to \{Q = i + 1\}$ since a customer arrives to the system. The interarrival time has an exponential distribution with parameter λ, hence the transition $\{Q = i\} \to \{Q = i + 1\}$ occurs with rate λ. We can illustrate these transitions using the following flow diagram, where we have chosen the case $i = 3$.



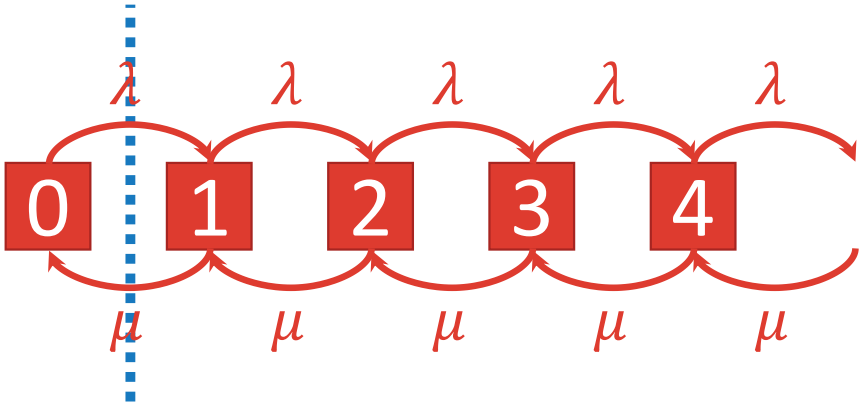
Doing this for all possible states we obtain the following flow diagram



We are going to compute the probabilities $p_k = \mathbb{P}(Q = k)$, for $k = 0, 1, \ldots$ using a flow conservation argument.

Flow Conservation Argument

The probability flux in any subset of states is equal to the probability flux out of that subset of states. Intuitively, this means that you enter a state just as many times as you leave a state.



Consider for example the set consisting of the state there are no customers in the system, that is $\{0\}$. Then the probability flux out of the set $\{0\}$ is λp_0 because we are in state 0 with probability p_0 and we leave it with rate λ. The probability flux into the set $\{0\}$ is equal to μp_1, since we can reach state 0 only from state 1, we are in state 1 with probability p_1 and the transition from state 1 to state 0 happens with rate μ. Then we get the first equation

$$\lambda p_0 = \mu p_1 \Rightarrow p_1 = \frac{\lambda}{\mu} p_0 = \rho p_0,$$

where $\rho = \frac{\lambda}{\mu}$. Suppose now that we consider the subset $\{0, 1\}$, then we obtain the equations

$$(\lambda + \mu) p_1 = \lambda p_0 + \mu p_2 \Rightarrow p_2 = \rho^2 p_0.$$

Working similarly we obtain the equations

$$(\lambda + \mu) p_i = \lambda p_{i-1} + \mu p_{i+1} \Rightarrow p_{i+1} = \rho^{i+1} p_0.$$

Hence it suffices to compute p_0, which denotes the probability that the system is empty. We know that the sum of all the probabilities has to be equal to one, hence

$$\sum_{i=0}^{\infty} p_i = 1.$$

Solving this equation yields

$$p_0 = 1 - \rho,$$

and we obtain the following result for the desired probabilities,

$$p_i = \rho^i(1-\rho), \quad i = 0, 1, \ldots. \tag{2.1.1}$$

Distribution of the number of customers

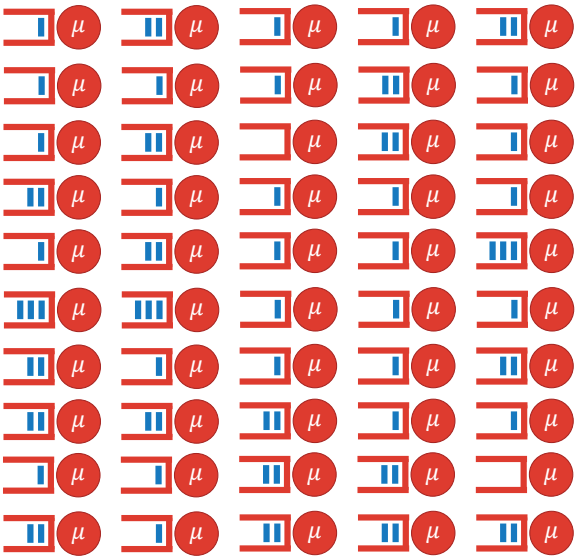
The random variable Q, denoting the number of customers in the queue when the system is in equilibrium, has a geometric distribution with parameter ρ.

2.1.3. Load balancing

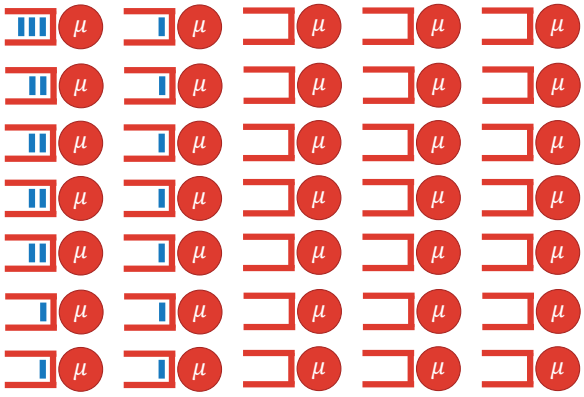
In many queueing systems jobs arrive at a central dispatcher and need to be sent to one of many servers. Typical examples are

- People arriving at the queues in a shop and need to be sent to one of the many cashiers;
- Texts arriving at a central server and need to be sent to one of many satellites;
- Cars arriving at a toll road and need to be sent to one of many booths;
- Data arriving at a dispatcher and need to be sent to one of many servers.

In such a system, consisting of thousands of servers, as in the case of data centres, we have a dispatcher who sends jobs to various queues. An illustration of such a system is given in the figure below.



First, let's consider what happens when the server would distribute jobs randomly. The dispatcher will randomly choose a server for every incoming job, with equal probability. This means that every server receives jobs with the same rate. Moreover, if you split a Poisson process into multiple streams, every stream is independent and also a Poisson process. So in this context, all of these servers will become independent M/M/1 queues. We observe that some queues are empty, other queues have one job, some queues more than one. If we order them, it would look like as in the figure below.



We study the fraction of queues that have i jobs at time 0, denote this quantity by $f_i(0)$. We have 24 queues which are empty, 10 queues which have one job, 5 queues which have two jobs and 1 queue which has three jobs. Hence we have that $f_0(0) = 0.6, f_1(0) = 0.25, f_2(0) = 0.125$ and $f_3(0) = 0.025$. At a later time customers will have arrived and will have departed, hence these fractions will have changed, an example is given in the figure below.

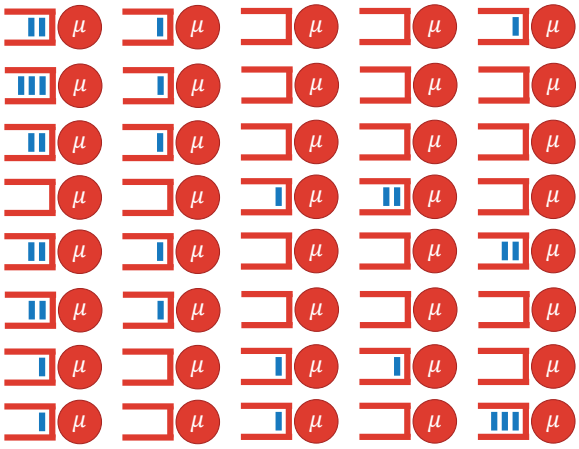
In this case we have $f_0(t) = 0.50, f_1(t) = 0.30, f_2(t) = 0.15$ and $f_3(t) = 0.05$. We want to compute how quickly these fractions change. Consider first $f_0(t)$, it can increase when a queue having one job becomes empty and it can decrease when a customer arrives to an empty system. The first event occurs with rate μ while the second event occurs with rate λ. This information shows how $f_0(t)$ changes over time, this can be quantified using the following differential equation

$$\frac{\mathrm{d}f_0(t)}{\mathrm{d}t} = -\lambda f_0(t) + \mu f_1(t). \tag{2.1.2}$$

Working similarly for $f_i(t)$ we obtain

$$\frac{\mathrm{d}f_i(t)}{\mathrm{d}t} = -\lambda(f_i(t) - f_{i-1}(t)) + \mu(f_{i+1}(t) - f_i(t)), \qquad \text{for } i = 0, 1, \ldots. \tag{2.1.3}$$

From this set of differential equations, you can once again find the equilibrium distribution. In order to find this, we set $\mathrm{d}f_i(t)/\mathrm{d}t = 0$, since the fraction should not change anymore



in equilibrium. After this, you obtain the exact same equation as for the M/M/1 queue, with $f_i(t) = p_i$, where p_i, for $i = 0, 1, \ldots$ was computed in (2.1.1).

Dispatching schemes

Popular job dispatching schemes are *join the shortest queue*, *random*, *round robin* and *power-of-d*.

In the *join the shortest queue* scheme the dispatcher sends a job to the server with the least amount of jobs, this is the most efficient job dispatching scheme but demands a lot of work from the dispatcher to keep track of all jobs in all servers. In the *random* scheme the dispatcher sends a job to a random server and in the *round robin* scheme the dispatcher sends jobs to servers in a cyclic way.

We will look in more detail the power-of-d balancing scheme. In this scheme when a job arrives you choose d servers at random and you dispatch the job at the one with the shortest queue. For simplicity let's have a look at the case $d = 2$. Denote by $g_i(t)$ the number of servers with queue length larger or equal than i. Then we have the following differential equation for $g_i(t)$,

$$\frac{\mathrm{d}g_i(t)}{\mathrm{d}t} = g_{i+1}(t) - g_i(t) + \lambda(g_{i-1}(t)^2 - g_i(t)^2). \tag{2.1.4}$$

If we consider the system to be in equilibrium, which means that the probability distribution describing its state does not change over time, we have then that the derivatives on the left-hand side above are equal to zero and the expressions on the right-hand side are independent of the time t. Hence we obtain the equations

$$g^*_{i+1} - g^*_i + \lambda({g^*_{i-1}}^2 - {g^*_i}^2) = 0, \tag{2.1.5}$$

which yields the solution

$$g_i^* \sim \lambda^{2^i-1}, \quad i \geq 0. \tag{2.1.6}$$

2.1.4. Load balancing with additional mechanics

So far, we have shown a very brief introduction to the mathematics that occur in simple queueing models. But it doesn't end here! There are many more models, and also many more ways to solve these type of models. We briefly name a couple of similar models.
First, we look at redundancy models. Every customer that joins the system, joins *multiple queues* at once. Whenever the customer is served by one of the servers, it will exit from all other queues. This sounds weird when you think of a supermarket, but it makes more sense in the context of data centres, where queries duplicate themselves and may be redundantly processed by multiple servers. In these models it is crucial that you duplicate sufficiently, but also not too much, as this will overload the system.
Finally, there are models where there are different types of customers and servers. Some customers may take more time, some servers may be slower, or maybe some customers can only be served by specific servers. And to make it worse, there are systems where jobs are not served in order of appearance. This can mean that the customer that arrived before you, can be processed first. This makes the analysis of such models much more complicated!

On the Network Pages

For an application of these techniques into data centres we refer to the article of Mark van der Boor, *The quest for a better Internet*, on the Network Pages, networkpages.nl/the-quest-for-a-better-internet/. In a relevant article, *Look it up on the Internet!" - How Web Search Works*, Nelly Litvak explains how web search works. Have a look at it on the Network Pages, networkpages.nl/look-it-up-on-the-internet-how-web-search-works/.

2.2. Networks in genetics – Understanding the evolution of genes

When you see the title you would maybe expect to read a biology book. However to describe the evolution of genes biologists actually use mathematics.

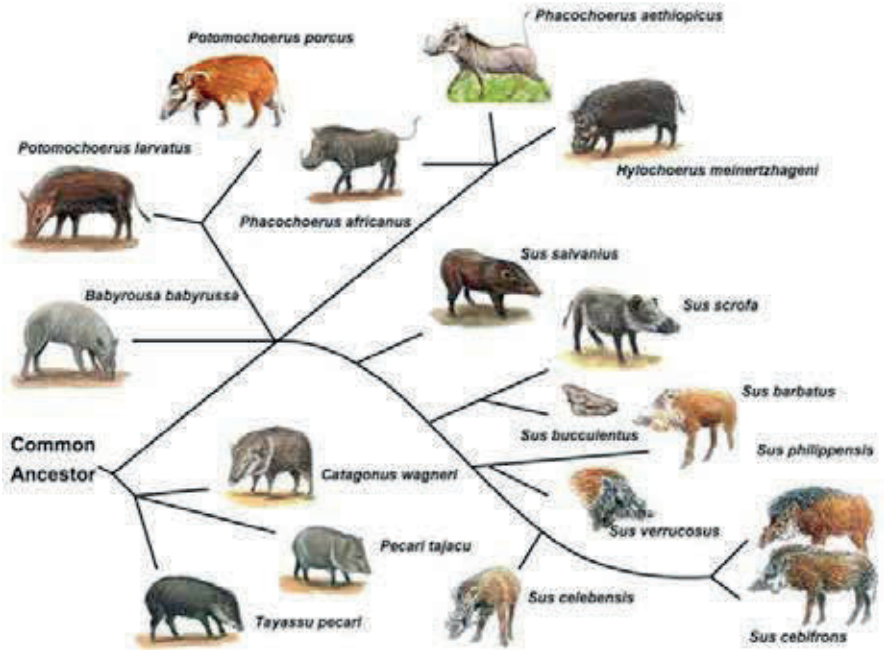


Figure 2.2.1. Evolution of pigs.

In Figure 2.2.1 [1] you see the evolution of pigs depicted. At certain moments a new pig species emerges. The DNA of this new pig species has different types of genes than his ancestors. With what probability does it happen that some mutation in the DNA occurs? And with what probability does the mutation in the DNA lead to a new species? To answer these questions you need probability theory.
In Figure 2.2.2[2] below you see another type of genetic evolution. This population of starfish lives at the Northeast Pacific and there are two types occurring in the population: purple starfish and orange starfish. You can ask yourself the question whether there will be always

[1] Taken from Chen K, Baxter T, Muir WM, Groenen MA, Schook LB. Genetic Resources, Genome Mapping and Evolutionary Genomics of the Pig (Sus scrofa). Int J Biol Sci 2007; 3(3):153-165. doi:10.7150/ijbs.3.153.

[2] Ochre sea stars (Pisaster ochraceus) taken at Ganges Harbour, Salt Spring Island, British Columbia, made by D. Gordon E. Robertson, for Wikipedia.

Figure 2.2.2. Population of starfish.

orange and purple starfish living together in this population. Or will at a certain time one type, orange or purple die out? And if the answer is yes, how long will it take before one of the two types dies out? In this lecture we will study the last question with the help of probability theory and a mathematical model, see Section 3.1.1 for an explanation of what a mathematical model is.

2.2.1. The Wright-Fisher model

The mathematical model we will use is called the Wright-Fisher model.

The Wright-Fisher model

In the Wright-Fisher model we have a population with two types of individuals. For example you can think about the starfish population.

The two types of starfish we consider are the purple starfish and the orange starfish. In the Wright-Fisher model you actually do not have to specify the species of your population, as long as there exist two different types you can use the model. So instead of starfish you can also think of a population with red and blue butterflies or a population of bacteria, where the bacteria have one of two types of a certain gene. To explain the Wright-Fisher model we use a two type population where the types are called red and blue. At each generation the population evolves according to the following two rules:

- Each individual chooses an individual (possibly itself) from the current generation and give birth to a child with the type of the chosen individual.
- The child replaces the parent.

These two rules that describe the evolution of the population are called *the dynamics* of our model. The production of offspring is called *resampling.* In Figure 2.2.3 you see an example of a population with five individuals and five generations. The generations are labelled by the time instances they live at.

From Figure 2.2.3 you see that the number of individuals in each generation is the same. This is because each child replaces its parent. At a first glance it seems not so natural that each parent gives birth to exactly one child and that the number of individuals in the population stays constant. However you can interpret the model in a different way. For example in Figure 2.2.3 the first individual on the left gets two children which inherit the blue type. The second individual gets no children. The third one gets one child, which has type blue, the fourth one gets two children that are red and the fifth one gets no children. This seems a more natural interpretation. The assumption that the total number of individuals in a population stays constant seems artificial if you have only five individuals, but in most cases the Wright-Fisher model is used to study large populations of individuals. As long as there are no drastic changes in the environment it is reasonable that the population stays

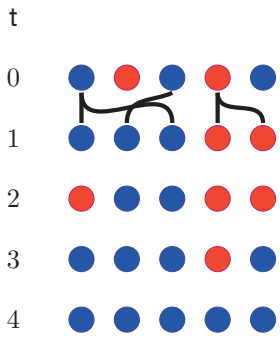


Figure 2.2.3. Wright-Fisher model with 5 individuals. An example of the outcome at the first 4 time steps is depicted. The black lines indicate which parent is chosen. So the first individual on the left has chosen the type of itself for its child and therefore stays blue. The second individual has chosen the third one and so on.

more or less constant. Therefore the model assumes that the total population stays constant over time.

2.2.2. Computing probabilities with the Wright-Fisher model

Mathematically we describe the Wright-Fisher model as follows:

$$X(t) = \text{number of red individuals at time } t. \tag{2.2.1}$$

For example in Figure 2.2.3 we have $X(0) = 2$, $X(1) = 2$, $X(2) = 3$, $X(3) = 1$ and $X(4) = 0$. Extending Figure 2.2.3 we should get $X(5) = 0$, since each individual at time 4 can only choose a blue individual. Therefore we say 0 is a *fixed point* of the dynamics. By a similar reasoning 5 is a fixed point of the dynamics, when our population consist of 5 individuals. Suppose we have 5 individuals in our population and in generation 0 there are 2 red individuals, so $X(0) = 2$. We want to compute the probability $\mathbb{P}(X(1) = 3)$. To do so we first introduce another random variable. For $0 \leq i \leq 5$ let

$$Y_i(t) = \begin{cases} 1 & \text{if at time } t \text{ individual } i \text{ is red} \\ 0 & \text{otherwise .} \end{cases} \tag{2.2.2}$$

To compute $\mathbb{P}(Y_1(1) = 1)$ recall that the first individual in generation 0 chooses at random another individual from the population. Since in generation 0 $X(0) = 2$, this is the same as drawing a ball from an urn containing 2 red balls and 3 blue balls. Therefore

$$\mathbb{P}(Y_1(1) = 1) = \frac{2}{5}.$$

Also the second individual chooses an individual randomly from the population. Hence also

$$\mathbb{P}(Y_2(1) = 1) = \frac{2}{5}$$

and in general we have

$$\mathbb{P}(Y_i(1) = 1) = \frac{2}{5}.$$

Therefore for all $1 \leq i \leq 5$, $Y_i(1)$ is a Bernoulli random variable with success probability $p = \frac{2}{5}$. The total number of red individuals in generation 1 can now be written as

$$X(1) = \sum_{i=1}^{5} Y_i(1).$$

Since the sum of n Bernoulli variables with success parameter p is distributed as $Bin(n, p)$, we refer to Section 2.1 for all necessary definitions and background knowledge, $X(1)$ is distributed as $Bin(5, \frac{2}{5})$. Hence

$$\mathbb{P}(X(1) = 3) = \binom{5}{3} \left(\frac{2}{5}\right)^3 \left(\frac{3}{5}\right)^2.$$

Since $\mathbb{E}[Y_i(1)] = 1 \cdot \frac{2}{5} + 0 \cdot \frac{3}{5}$ for all $1 \leq i \leq 5$, we can easily compute $\mathbb{E}[X(1)]$ with the help of the sum formula for expectations. Namely

$$\mathbb{E}[X(1)] = \mathbb{E}[\sum_{i=1}^{5} Y_i(1)] = \sum_{i=1}^{5} \mathbb{E}[Y_i(1)] = 5 \cdot \frac{2}{5} = 2. \tag{2.2.3}$$

Therefore if we start with 2 red individuals at timestep 0, we actually also expect 2 red individuals in timestep 1. So the expected value of the number of red individuals has not changed from time 0 to time 1.
Suppose we know $X(1) = 2$, then

$$\mathbb{P}(Y_i(2) = 1 | X(1) = 2) = \frac{2}{5}$$

for all individuals $1 \leq i \leq 5$. Hence given that $X(1) = 2$ we have that $Y_i(2)$ is distributed as a Bernoulli variable with success probability $\frac{2}{5}$ and therefore $X(2)$ is distributed $Bin(5, \frac{2}{5})$. For example

$$\mathbb{P}(X(2) = 4) = \binom{5}{4} \left(\frac{2}{5}\right)^4 \left(\frac{3}{5}\right).$$

In Exercise 3.3.1 you will compute more probabilities using the the Wright-Fisher model.

2.2.3. Genetic Diversity

With the Wright-Fisher model you can examine *genetic diversity*.

Genetic diversity

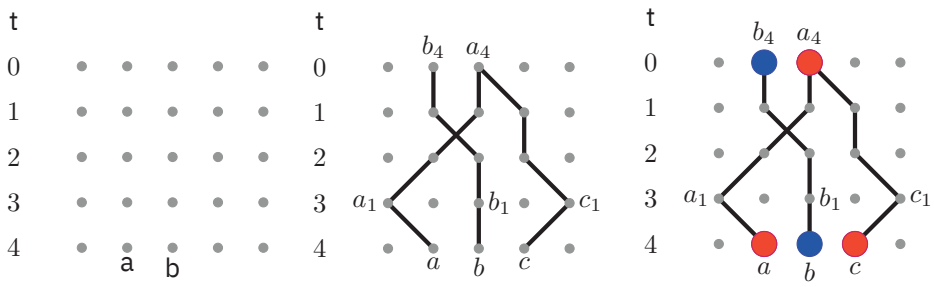
We say that there is *genetic diversity* in a population if there are both red and blue individuals in the population.

To compute the probability that there is genetic diversity in a certain generation, we pick two individuals from that generation at random. Subsequently we compute the probability that the two individuals are of the same type with the following procedure, which is also illustrated in Figure 2.2.4 below for a population of five individuals.
In Figure 2.2.4a two randomly drawn individuals, denoted by a and b are shown. In Figure 2.2.4b a line from individual a is drawn to its parent in third generation ($t = 3$), which we call a_1 and the same is done for b, so that a parent b_1 is found. From a_1 a line is drawn to its parent in the second generation ($t = 2$) and this continues until the parent of a at time $t = 0$ is found, which is called a_4. In same way the parent b_4 of b is found. The paths from a to a_4 and b to b_4 are called *the (ancestral) lineages* of respectively a and b. The parent a_4 of a in generation 0 is called the ancestor of a and similarly b_4 is the ancestor of b. Since in each time step an individual inherits its type from its parent, the type of individual a is the same as the type of its ancestor a_4. Similarly the type of individual b is the type of its ancestor b_4. Note that the paths drawn to the ancestors a_4 and b_4 in generation 0 are only one of all the possible paths from a to a_4 and b to b_4.
In Figure 2.2.4b the ancestral lineage for a third individual c are also drawn. In this instance the lineages of a and c have the same ancestor in generation 0 and therefore they have the same type. If you know that two individuals have the same ancestor, you know that they have the same type. If they have don't have the same ancestor, the probability of having the same type depends on the probability of whether their ancestors in first generation are of the same type. The precise computation is done step by step in Exercise 3.3.3. The procedure above gives us a direct criterion to check whether there is genetic diversity in the population or not, namely:

Criterion for genetic diversity

If all individuals in a generation have the same ancestor in the first generation, the genetic diversity is lost.

In Exercise 3.3.3 we will compute step by step the probability that genetic diversity is lost in a certain generation.



(a) Pick two individuals a and b at random

(b) Determine their ancestors in generation 0

(c) Determine the type of the ancestors in generation 0

Figure 2.2.4. Procedure to check genetic diversity

2.2.4. Scaling limit

Suppose we want to compare two populations evolving according to the Wright-Fisher model with different sizes. Let N_1 denote the number of individuals in population 1 and N_2 denote the number of individuals in population 2. To compare the two populations it is convenient to look at the fraction of red individuals, instead of the absolute number of red individuals, i.e.

$$\tilde{X}_1(t) = \frac{X_1(t)}{N_1} \quad \text{and} \tag{2.2.4}$$

$$\tilde{X}_2(t) = \frac{X_2(t)}{N_2}. \tag{2.2.5}$$

The Wright-Fisher model is often used to study the genetic composition of a population of bacteria. In that case the number of individuals in the population is typically very large. Also then it makes sense to look at $\tilde{X}(t) = \frac{X(t)}{N}$. You can imagine that for a large population it will take much longer for one of the two types to die out, then for a population with only a few individuals. Therefore if you want to study a large population it makes sense to let time pass by faster in order to see the population change. You can compare this to a growing plant. If you watch the plant, you will not see it grow. If you film the plant for a week and speed up the video afterwards, you are able to see the plant grow. Therefore, for large population with N individuals, we are interested in the quantity

$$\bar{X}(t) = \frac{X(Nt)}{N}. \tag{2.2.6}$$

So $\bar{X}(t)$ is the fraction of red individuals when time is speeded up by a factor N, the size of the population. This means that a timestep which has before speeding up size 1, now has size $\frac{1}{N}$

It turns out that when N grows very large the process $\bar{X}(t)$ can be described by the following *differential equation*

$$\mathrm{d}\bar{X}(t) = \sqrt{\bar{X}(t)(1-\bar{X}(t))}\mathrm{d}W(t). \tag{2.2.7}$$

Here $\mathrm{d}\bar{X}(t)$ tells us how much the process changes when time is changed a little bit. $W(t)$ is random process called a Brownian motion that takes random values. A lot of mathematics is needed to understand equation (2.2.7), but we are still able to grab some things. Equation (2.2.7) tells us, that the fraction of red individuals change over time like a random process changes over small times, this is $\mathrm{d}W(t)$, multiplied by a factor $\sqrt{\bar{X}(t)(1-\bar{X}(t))}$. This implies that if there are only red individuals in the population so $X(t) = 1$, then $\sqrt{\bar{X}(t)(1-\bar{X}(t))} = 0$ and so the process does no longer change over time. The same holds for $X(t) = 0$. We say that 0 and 1 are the fixed values of the *differential equation*. The process $\bar{X}(t)$ is called the scaling limit.

2.2.5. Wright-Fisher model on the hierarchical lattice

The Wright-Fisher model can be extended in multiple ways. You can think for example about mutations or selections of a fitter type or migration between different colonies. In this section we would look at the last extension where migration takes place on the hierarchical lattice, see Section 2.2 for the necessary background knowledge. The colonies are then placed at the sites of the 0-layer of the hierarchical lattice. The hierarchical lattice is used often in ecology because it orders the colonies in natural way. Denoting the site of a colony by the sequence given in (1.3.1), we can interpret each colony as a house, all the colonies with the same second entry form together a street, all colonies with the same third entry form a neighbourhood, the neighbourhoods form cities, the cities form countries and so on. The individuals can migrate between colonies. The migration is incorporated in the Wright-Fisher model in the following way.

Migration in the Wright-Fisher model

At each timestep an individual at a site η chooses with probability $a(\eta, \xi)$ an individual from site ξ and adopts its type.

Recall the distance $d(\eta, \xi)$ on the hierarchical lattice defined in (1.3.2). The probability $a(\eta, \xi)$ to migrate from site $\eta \in \Omega_M$ to site $\xi \in \Omega_M$ is defined as follows

$$a(\eta, \xi) = C \sum_{\substack{k \in \Omega^M \\ k \geq d(\eta,\xi)}} \frac{1}{M^{k-1}} \frac{1}{M^k}, \tag{2.2.8}$$

where M is the order of the hierarchical lattice and C is a constant to make sure that $\sum_{\xi \in \Omega_M} a(\eta, \xi) = 1$. This migration probability should be interpreted as follows. With probability $\frac{C}{M^{k-1}}$ an individual chooses a space horizon, (i.e. house, street, neighbourhood,

city, ...) from which it subsequently picks a colony at random. Since a space horizon k contains M^k colonies, the probability to pick at random the colony ξ from this space horizon is exactly $\frac{1}{M^k}$.
If M is large $\frac{1}{M} >> \frac{1}{M^2} >> \frac{1}{M^3} >> \cdots$, where the notation $>>$ means that the term on the left is much larger than the term on the right. Hence we see that migration in the street happens far more often than migration in the neighbourhood, which in turn happens far more often than migration in the city. To study how a street, a house or a city evolves we study the so called blocks. The block of level k is defined as follows:

$$X_\eta^{[k]}(t) = \frac{1}{M^k} \sum_{\substack{\xi \in \Omega_d \\ d(\eta,\xi) \leq k}} \bar{X}_\eta(M^k t), \tag{2.2.9}$$

so it averages the fraction of red individuals in all single colonies within a certain space horizon, while time is speeded up by a factor M^k. The time is speeded up to see the migration occur between blocks of different sizes. This is called space time scaling. At this point it is to difficult to analyse the dynamics of the Wright-Fisher model with migration in detail, but the pictures below illustrate how you can look at different timescales to see the system change at the corresponding space scale.

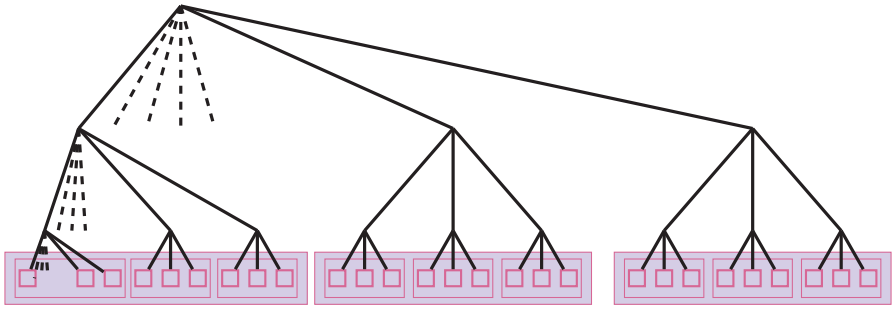


Figure 2.2.5. The hierarchical lattice Ω_M at time 0. The dots indicate that there are M lineages instead of three, but for clearness of the picture there are only three depicted. The purple colour indicates there is a mixture between red and blue individuals.

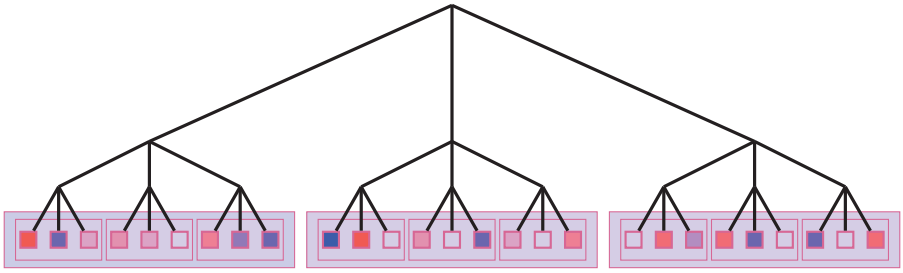


Figure 2.2.6. The hierarchical lattice Ω_M at time t. The single colonies start their evolution, but bigger blocks remain in there original state.

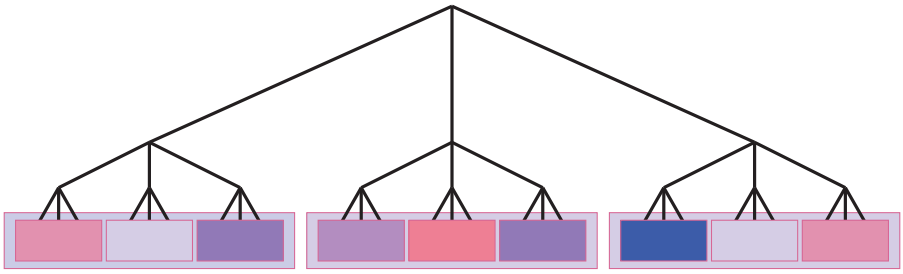


Figure 2.2.7. The hierarchical lattice Ω_M at time Mt, the 1-blocks start evolving

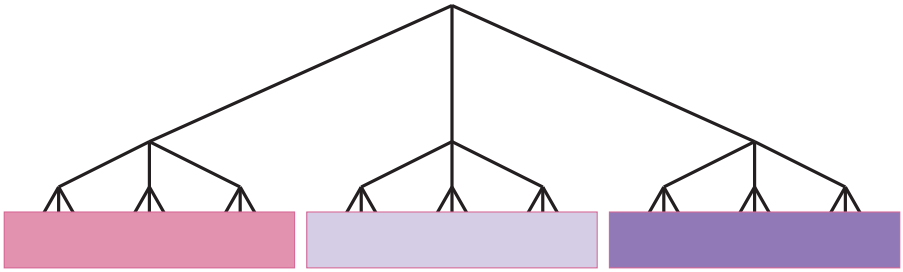


Figure 2.2.8. The hierarchical lattice Ω_M at time M^2t, the 2-blocks start evolving.

On the Network Pages

In her article, *The seed bank, an unseen storage of genetic diversity*, on the Network Pages, Margriet Oomen goes deeper into the mysteries of evolution and genetics. Have a look at networkpages.nl/the-seed-bank-an-unseen-storage-of-genetic-diversity.

CHAPTER 3

Exercises

This chapter contains exercises on the theory presented in chapters 1 and 2.

3.1. Exercises on probability theory

3.1.1. Conditional probabilities and expectations

A conditional probability is denoted by $\mathbb{P}(A|B)$, which corresponds to

the probability of A happening, given that B happens.

Let's look at a few simple examples. We denote by X the random variable that represents the number that you roll with a six-sided die.

(1) What is the probability that you roll a 6 with a six-sided die? In a formula: $\mathbb{P}(X = 6)$.
(2) What is the probability that you roll a 6, given that you roll at least a 4; $\mathbb{P}(X = 6|X \geq 4)$?
(3) You can use the following formula to compute conditional probabilities:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \text{ and } B)}{\mathbb{P}(B)}. \tag{3.1.1}$$

Check that this formula works by solving the second question again, but now with the formula.
(4) Similarly to probabilities, we can also look at expectations. What is the expected number you roll with a six-sided die? In formulas: $\mathbb{E}(X)$.
(5) What is the expected number that you roll, given that you roll at least a 4; $\mathbb{E}(X|X \geq 4)$?

3.1.2. The exponential distribution

The exponential distribution is defined in the following way. Suppose that X is exponentially distributed with parameter λ. Then $\mathbb{P}(X < t) = 1 - e^{-\lambda t}$.

(1) Calculate $\mathbb{P}(X \geq t)$.
(2) Calculate $\mathbb{P}(1 < X < 2)$.
(3) Calculate the expectation of the exponential distribution with the following formula:

$$\mathbb{E}(X) = \int_0^\infty \mathbb{P}(X \geq t)dt.$$

(4) Use Equation (3.1.1) to prove the memoryless property of the exponential distribution:

$$\mathbb{P}(X > t + u|X > t) = \mathbb{P}(X > u).$$

3.2. Exercises on queueing theory

3.2.1. Mean queue length

We introduce $\rho = \lambda/\mu$ to make the calculations easier. In the $M|M|1$ queue we found that the probability of having i jobs in the system, in equilibrium, equals

$$p_i = (1-\rho)\rho^i.$$

(1) Of course, the sum of all these probabilities should sum up to 1. Prove that $\sum_{i=0}^{\infty} p_i = 1$.

(2) We can calculate the mean queue length using these probabilities;

$$\mathbb{E}(L) = \sum_{i=0}^{\infty} i p_i = \sum_{i=0}^{\infty} i(1-\rho)\rho^i.$$

Calculate $\mathbb{E}(L)$.

3.2.2. Extension of the single-server queue

In the lecture we drew the transition diagram and calculated the equilibrium probabilities of the $M|M|1$ queue, which is a system where 1 job can be served at a time. In this set of questions, we will consider three extensions.

(1) The $M|M|c$ queue is an extension of this model, where up to c jobs can get service simultaneously. Draw the transition diagram of the $M|M|c$ queue.

Hint

Suppose two jobs are getting service at the same time. The rate at which servers move from having 2 to 1 jobs, is equal to $2 \cdot \mu$.

Calculate the equilibrium probabilities of the $M|M|c$ queue.

(2) In the $M|M|1|k$ queue, only one job receives service at a time. The k in the name denotes that there are finitely many spots to wait in the queue. At any moment, there can be at most k jobs in this system. Whenever a job arrives and the system is full, it will be blocked and it will leave forever. Draw the transition diagram, calculate the equilibrium probabilities, and find the blocking probability; the probability that an arbitrary job will be blocked.

(3) The $M|M|c|k$ model is a mix of the $M|M|1|k$ and the $M|M|c$. In this system there are c servers, hence c jobs can receive service simultaneously, and at most k jobs can reside in the queue. Can you find the transition diagram, equilibrium probabilities and blocking probability?

3.2.3. Additional questions

(1) For the 'random' dispatching, we showed differential equations, involving the fraction of servers that have i jobs in them at time t; $f_i(t)$. Can you explain what these formulas represent?

$$\frac{df_0(t)}{dt} = -\lambda f_0(t) + \mu f_1(t)$$

and

$$\frac{df_i(t)}{dt} = \lambda(f_{i-1}(t) - f_i(t)) + \mu(f_{i+1}(t) - f_i(t)), i \geq 1.$$

(2) For Power-of-2, we showed differential equations, involving the fraction of servers that have *at least* i jobs in them at time t; $g_i(t)$. Can you explain what these formulas represent?

$$\frac{dg_i(t)}{dt} = g_{i+1}(t) - g_i(t) + \lambda(g_{i-1}(t)^2 - g_i(t)^2), i \geq 1.$$

(3) Can you think of a model where Join-the-Shortest-Queue is *not* smart to use?

(4) In the presentation, you saw several load balancing algorithms like Join-the-Shortest-Queue and power-of-d. Can you think of your own load balancing algorithm?

3.3. Exercises on mathematical genetics

3.3.1. Wright–Fisher model

Consider the Wright-Fisher model from Section 2.2.1.

(1) Suppose that the population has 6 individuals and that $X(0) = 4$, what is $\mathbb{P}(X(1) = 3)$? And what is $\mathbb{E}[X(1)]$?

(2) Suppose that the population has N individuals and we know $X_0 = K$ for some $1 \leq K \leq N$. Let $0 \leq l \leq N$, can you show that $\mathbb{P}(X(1) = l) = \binom{N}{l}(\frac{K}{N})^l(\frac{N-K}{N})^{N-l}$ and $\mathbb{E}[X(1)] = K$?

(3) Note that $\mathbb{E}[X(1)] = X_0$, can you guess what $\mathbb{E}[X(n)]$ will be? (Hint: suppose you know X_{n-1}, can you express the probability $\mathbb{P}(X_n = l)$ in terms of X_{n-1}? Try to compute $\mathbb{E}[X(n)|X(n-1)] = \sum_{l=0}^{N} l \cdot \mathbb{P}(X(n) = l|X(n-1))$, this is the expectation of $X(n)$ if you know what $X(n-1)$ is.)

(4) Suppose there exists a generation m such that $X_m = 0$ or $X_m = N$. What do we know for X_n with $n > m$, so the generations after generation m?

3.3.2. Hierarchical lattice Ω_4

Answer the question 1 and 2 by drawing it in Figure 3.3.1.

(1) Give the sequence which indicates the positions of ξ and η.

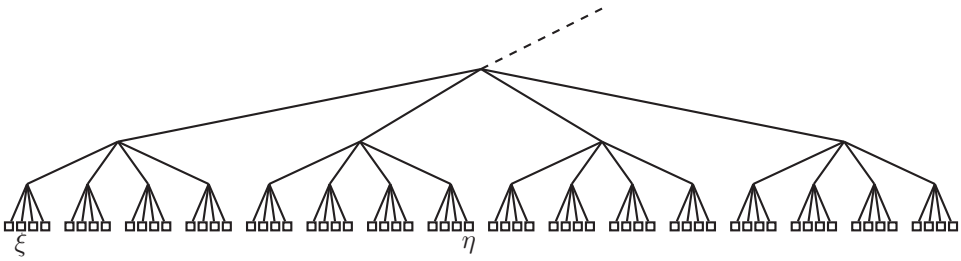


Figure 3.3.1. Hierarchical lattice of order 4, Ω_4.

Hint

We start enumerating from 0 and from left to right.

(2) Draw a block of all individuals at distance 1 from ξ and at distance 2 and at distance 3.
(3) What is the distance between ξ and η?

3.3.3. Loss of genetic diversity

In this question we compute what the probability is that two individuals are of the same type. We consider the Wright-Fisher model with N individuals. The number of red individuals at time 0 is x_0, $1 \leq x_0 \leq N$.

(1) Show that two individuals a and b chosen at random from the population at time 1 have the same parent equals $\frac{1}{N}$. Show that the same holds for the probability that two individuals chosen at random from the population at time n have the same parent in generation $n-1$.
(2) From (1) it follows that the probability two individuals in generation n do not have the same parent equals $(1-\frac{1}{N})$. Show that the probability that two individuals in generation n don't have the same ancestor in generation 0 equals $(1-\frac{1}{N})^n$.
(3) We know that the initial number of red individuals $X_0 = x_0$. Show that the probability that the two individuals a and b in generation n are of a different type equals $(1-\frac{1}{N})^n \frac{2x_0(N-x_0)}{N(N-1)}$.What is the probability that the two individuals a and b in generation n are of the same type?
(4) What happens with the probability computed in (3) if n gets larger?
(5) Suppose we know that all the N individuals in our population have the same ancestor. What do you think is the probability that this ancestor is RED? You may assume that each individual in the 0 generation is equally likely to be the ancestor of the population.
(6) If we have a single colony do you think that one type always get extinct? Can you think of extensions of the model such that types get less often extinct?

CHAPTER 4

Solutions to the exercises

This chapter contains the solutions to some of the exercises in chapter 3.

4.1. Probability theory

4.1.1. Conditional probabilities and expectations

(1)
$$\mathbb{P}(X = 6) = \mathbb{P}(\text{you get a 6 when rolling a six-sided die}) = \frac{1}{6},$$
since it is equally probable to obtain any of the six sides.

(2) This is a conditional probability. You don't know exactly what the outcome is but you know that it is at least 4. This means that the die number is either a 4 or a 5 or a 6. Yes now you have three possible outcomes, given the condition, not six. All three are equally probable, hence the desired probability is equal to
$$\mathbb{P}(X = 6|X \geq 4) = \frac{1}{3}.$$

(3)
$$\mathbb{P}(X = 6|X \geq 4) = \frac{\mathbb{P}(\{X = 6\} \text{ and } \{X \geq 4\})}{\mathbb{P}(X \geq 4)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}. \tag{4.1.1}$$

(4)
$$\mathbb{E}(X) = \sum_{i=1}^{6} i\mathbb{P}(X = i) = \frac{1}{6}\sum_{i=1}^{6} = 3.$$

(5)
$$\mathbb{E}(X|X \geq 4) = \sum_{i=1}^{6} i\mathbb{P}(X = i|X \geq 4) = 5.$$

4.1.2. The exponential distribution

(1)
$$\mathbb{P}(X \geq t) = 1 - \mathbb{P}(X < t) = e^{-\lambda t}.$$

(2)
$$\mathbb{P}(1 < X < 2) = \mathbb{P}(X < 2) - \mathbb{P}(X < 1) = e^{-\lambda} - e^{-2\lambda}.$$

(3)
$$\mathbb{E}(X) = \frac{1}{\lambda}.$$

(4)
$$\mathbb{P}(X > t + u \text{ and } X > t) = \mathbb{P}(X > t + u),$$
because if $X > t + u$ then it will also happen that $X > t$. The rest follows by doing one more computation.

4.2. Queueing theory

4.2.1. Mean queue length

(1)

$$\sum_{i=0}^{\infty} p_i = \sum_{i=0}^{\infty} (1-\rho)\rho^i = (1-\rho)\sum_{i=0}^{\infty} \rho^i.$$

Geometric sum

For the geometric sum we have that

$$\sum_{i=0}^{n} \omega^i = \frac{1-\omega^{n+1}}{1-\omega}.$$

Hence we have that

$$\sum_{i=0}^{\infty} \omega^i = \lim_{n\to\infty} \sum_{i=0}^{n} \omega^i = \lim_{n\to\infty} \left(\frac{1-\omega^{n+1}}{1-\omega} \right),$$

and hence for $\omega \in (0,1)$

$$\sum_{i=0}^{\infty} \omega^i = \frac{1}{1-\omega}.$$

Using this result the answer follows.

(2)

$$\sum_{i=0}^{\infty} i(1-\rho)\rho^i = \sum_{i=1}^{\infty} i(1-\rho)\rho^i = \rho(1-\rho)\sum_{i=1}^{\infty} i\rho^{i-1}$$
$$= \rho(1-\rho)\left(\sum_{i=0}^{\infty} \rho^i\right)' = \rho(1-\rho)\left(\frac{1}{1-\rho}\right)' = \frac{\rho}{1-\rho}.$$

4.2.2. Additional questions

(1) In order to understand these differential equations we have to see how the quantity $f_0(t)$ changes in time. First of all a reminder, $f_0(t)$ represents the fraction of queues with 0 customers. Hence this fraction becomes smaller when the number of queues with no customers decreases, this happens only when a customer arrives to one of these queues. Moreover, customers arrive to the system with rate λ. This is why the term $-\lambda f_0(t)$ appears on the right-hand side. Now it remains to see how $f_0(t)$ can become larger, this happens when a customer departs from a queue with only one cus-

tomer and the fraction of queues with one customer is equal to $f_1(t)$. Moreover, customers are served with rate μ. This is why the term $\mu f_1(t)$ appears on the right-hand side of the equation. Let's move on to the equation for $i \geq 1$. The same argument holds but now we have two more scenarios. The fraction of queues with i customers can increase also because a customer arrived to a queue with $i-1$ customers, that is why the term $\lambda f_{i-1}(t)$ appears. And it can decrease because a customer in one of the queues with i customers is served and leaves the queue, that is why the term $-\mu f_i(t)$ appears.

4.3. Mathematical genetics

4.3.1. The Wright–Fischer model

(1) Note $X(1)$ is now distributed as $Bin(6, \frac{4}{6})$ and hence

$$\mathbb{P}(X(1)=3) = \binom{6}{3}\left(\frac{4}{6}\right)^3\left(\frac{2}{6}\right)^3.$$

In this case $\mathbb{P}(Y_i(1)=1) = \frac{4}{6}$ Hence

$$\mathbb{E}[X(1)] = \mathbb{E}[\sum_{i=1}^{6} Y_i(1)] = \sum_{i=1}^{6}\mathbb{E}[Y_i(1)] = 4.$$

Or you can use immediately that for a random variable X that is distributed $Bin(n,p)$ $\mathbb{E}[X] = n\cdot p$. So in our case $\mathbb{E}[X] = 6\cdot\frac{4}{6} = 4$.

(2) In this case X is $Bin(N, \frac{K}{N})$ distributed, hence by the above rules $\mathbb{P}(X(1) = l) = \binom{N}{l}(\frac{K}{N})^l(\frac{N-K}{N})^{N-l}$ and $\mathbb{E}[X_1] = K$.

(3)

$$\mathbb{P}(X(n)=l|X(n-1)) = \binom{N}{l}\left(\frac{X(n-1)}{N}\right)^l\left(\frac{X(n-1)}{N}\right)^{n-l},$$

so we see that $X(n)$ is $Bin(N, \frac{X(n-1)}{N})$ distributed. Hence $\mathbb{E}[X(n)|X(n-1)] = X(n-1)$, so note the value $\mathbb{E}[X(n)|X(n-1)]$ depends on the value $X(n-1)$ assumes. Taking expectations once more we see

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X(n)|X(n-1)]] &= \sum_{l=0}^{N}\mathbb{P}(X(n-1)=l)\cdot\mathbb{E}[X(n)|X(n-1)=l]\\ &= \sum_{l=0}^{N}\mathbb{P}(X(n-1)=l)\cdot l = \mathbb{E}[X(n-1)].\end{aligned}$$

Since the above result holds for all n, we see $\mathbb{E}[X(0)] = \mathbb{E}[X(1)] = \mathbb{E}[X(2)] = ... = \mathbb{E}[X(n)]$. But we actually knew that $X(0) = K = \mathbb{E}[X(1)]$. Therefore we see that

$\mathbb{E}[X(n)] = X(0)$ for all n, so the number of red individuals we expect in a certain generation stays constant over time.

(4) If $X(m) = 0$, then all the individuals are blue, hence all individuals will choose a blue individual and therefore all children will be blue so $X(m+1) = 0$. Therefore for all $n > m$ $X(m+1) = 0$. The same reasoning holds for $X(m) = N$.

4.3.2. Hierarchical lattice Ω_4

(1) $\xi = (1, 0, 0, 0, \cdots)$, $\eta = (3, 3, 1, 0, \cdots)$. To find the answer, number in each layer the branches of one node from left to right by $\{0, 1, 2, 3\}$.

(2) Draw a block of all individuals at distance 1 from ξ and at distance 2 and at distance 3.

(3) $d(\xi, \eta) = 3$, since they agree from the fourth coordinate onwards.

4.3.3. Loss of genetic diversity

(1) If we number the individuals at time 0 from 1 to N then

$$\mathbb{P}(a \text{ descends from individual } l) = \frac{1}{N}$$

for all $1 \leq l \leq N$. Similarly $\mathbb{P}(b$ descends form individual $l) = \frac{1}{N}$. Since a and b choose there parent independently, $\mathbb{P}(a$ and b descends form individual $l) = \left(\frac{1}{N}\right)^2$. Since the only requirement is that a and b descend from the same individual we have $\mathbb{P}(a$ and $b$descends form the same individual$) = \sum_{l=0}^{N} \left(\frac{1}{N}\right)^2 = \frac{1}{N}$. Equivalently one can argue that a can choose any parent. Then b has exactly to choose this parent, which happen with probability $\frac{1}{N}$. Hence it follows that $\mathbb{P}(a$ and b have the same parent $) = \frac{1}{N}$.

(2) To have not the same ancestor n generations back, the "parents" of a and b have to choose n times a different ancestor, which each time happens with probability $(1 - \frac{1}{N})$. Hence the probability that two individuals in generation n have not the same ancestor in generation 0 equals $(1 - \frac{1}{N})^n$. Note that as soon as in some generation the "parents" of a and b choose the same line, the ancestral path will be the same from this point on, backwards in time. See the ancestral lines of b and c, in Figure 2.2.4b.

(3) We define two events A and B. Let

$$A = \{a \text{ and } b \text{ have same ancestor in generation } 0\}$$

and

$$B = \{a \text{ and } b \text{ have different ancestor in generation } 0\}.$$

Then $\mathbb{P}(A$ or $B) = 1$ and from part B we know that $\mathbb{P}(B) = (1 - \frac{1}{N})^n$. We also know that

$$\mathbb{P}(a \text{ and } b \text{ are of different types}|A) = 0,$$

since if a and b have the same ancestor, they inherit both the type of this ancestor and therefore a and b are of the same type. And $\mathbb{P}(a \text{ and } b \text{ are of different types}|B) = 2 \cdot \frac{x_0}{N} \cdot \frac{N-x_0}{N}$, since the probability that two randomly drawn individuals at time 0 are of different type is the same as drawing two balls without replacement from an urn with x_0 red balls and $N - x_0$ blue balls and having either the outcome red, blue or the outcome blue, red.

$$\begin{aligned}
\mathbb{P}(a \text{ and } b \text{ are of different types}) &= \mathbb{P}(a \text{ and } b \text{ are of different types}|A)\mathbb{P}(A) \\
&\quad + \mathbb{P}(a \text{ and } b \text{ are of different types}|B)\mathbb{P}(B) \\
&= \mathbb{P}(a \text{ and } b \text{ are of different types}|B)\mathbb{P}(B) \\
&= \frac{2x_0(N-x_0)}{N(N-1)} \cdot (1 - \frac{1}{N})^n
\end{aligned}$$

(4) If n gets larger, $(1 - \frac{1}{N})^n$ tends to 0. Since $0 \leq \frac{2x_0(N-x_0)}{N(N-1)} \leq 1$ is a constant independent of n, $\mathbb{P}(a \text{ and } b \text{ are of different types})$ tends to 0.

(5) Since each individual in the population is equally likely to be the ancestor, the probability that this ancestor is red equals $\frac{x_0}{N}$.

(6) Indeed one of the types gets extinct. To argue this answer you could reason that the amount of red or blue individuals only stops changing when every one is red or blue. Since in each timestep this could happen with some positive probability, this will happen after long enough time. In each evolution step of the population we have a probability of $(\frac{1}{N})^{N-1}$ that all the individuals choose the same parent. Therefore you know that the probability that every one has the same color at the next time is at least $(\frac{1}{N})^{N-1} > 0$, since also choosing different parents with the same color can result in a loss of genetic diversity. Also in part C we saw that the probability for two randomly drawn individuals to be of different type tends closer to 0 if we look at larger times. Once this probability becomes 0 we are sure all pairs are of the same type.

NET
WORKS

by address: University of Amsterdam
Faculty of Science - Korteweg-de Vries
Institute
PO Box 94248
1090 GE Amsterdam
the Netherlands
+31 (0)20 525 6499
info@thenetworkcenter.nl
www.thenetworkcenter.nl

design www.dezagerij-ontwerp.nl • **lay-out** Vincent Schmeits

WWW.THENETWORKCENTER.NL