

Lesson 1. Introduction to Simulation

1 What is simulation?

- **Simulation** is the imitation of a real-world system in order to obtain data that can be used to evaluate and improve the system's performance
- In this course, we will focus on simulations that are
 - **stochastic**: some aspects of the system are modeled using random variables
 - **discrete-event**: state of the system changes at discrete points in time triggered by **events**, e.g. the arrival of a customer, the completion of an activity

2 Why simulate?

- Real-world trial-and-error approaches are expensive, time consuming and disruptive
- Complex systems are often resistant to analytical models and solutions (e.g. limitations of mathematical programming, queueing theory)

3 The Fantastic Dan Problem

Customers visit the neighborhood hair stylist Fantastic Dan for haircuts. The customer interarrival time is exponentially distributed with mean 20 minutes. Each haircut takes Fantastic Dan anywhere from 15 to 25 minutes, uniformly distributed.

- Questions we might be interested in:
 - On average, how long does a customer spend at the hair stylist?
 - On average, how many customers are waiting for a haircut?
- We can answer these questions by simulating the arrival and service of customers
- To perform this simulation, we need:
 - time of arrival of each customer
 - how long it takes to serve each customer
- Obtaining these requires **sampling** from the above probability distributions
- We will discuss the mathematics behind sampling later in the course
- For now, let's assume we have a sampling oracle for these distributions

3.1 Simulating the first 5 customers

- Fantastic Dan opens his shop at time 0
- The interarrival times of the first 5 customers are 23, 5, 10, 79, 13
- The service times of the first 5 customers are 17, 21, 19, 22, 19

Customer (a)	Interarrival time (b)	Arrival time (c)	Start service time (d)	Service time (e)	Departure time (f)	Total time in shop (g)

- Arrival time c = time when customer arrives at shop

- Start service time d = time when customer starts service

- Departure time f = time when customer leaves =

- Total time at shop g =

- Does arrival time c always equal start service time d ?

- What assumptions are we making with these calculations?

4 Performance measures

- We simulate a system in order to understand how the system performs
- Some common **performance measures**, especially for systems with queues:
 - average waiting time and average delay
 - server utilization
 - time average number of customers

4.1 Average waiting time and average delay

- The **average waiting time** w is the average time a customer spends in the system from arrival to departure
- Offers a view of performance from the customer's perspective
- For the Fantastic Dan simulation above, the average waiting time is

- We can similarly compute the **average delay** w_q : the average time a customer spends in the queue

4.2 Server utilization

- The **server utilization** ρ is the proportion of time that the server is busy serving a customer:

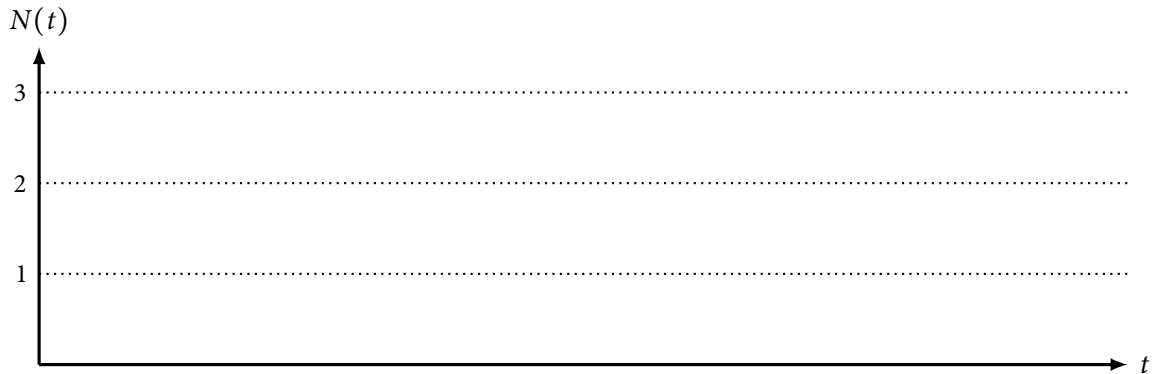
- Offers a view of performance from the system's perspective
- For the Fantastic Dan simulation above, the server utilization from the time the shop opens until the time the 5th customer departs is

4.3 Time average number of customers

- The **time average number of customers** ℓ in time interval $[0, T]$ is the average number of customers in the system at any time in $[0, T]$
- Offers another view of performance from the system's perspective
- Mathematically:
 - Let $N(t)$ be the number of customers in the system at time t

⇒

- For the Fantastic Dan simulation above, let's graph $N(t)$ for $0 \leq t \leq 158$



- Then the time average number of customers in time interval $[0, 158]$ is

- Alternatively, suppose $N(t)$ changes value at $t_1, t_2, \dots, t_{m-1}, t_m$
- For example, for the Fantastic Dan simulation above

- Then the time average number of customers can be computed as