SIMULATION & MODELLING

CSE-504

1

Md. Masudur Rahman, Dept. of CSE, UGV

SIMULATION AND MODELLING

Course Code:	CSE-504	Credits:	03
		CIE Marks:	90
Exam Hours:	03	SEE Marks:	60

Course Learning Outcome (CLOs): After Completing this course successfully, the student will be able to…

CLOs	Description
CLO1	Understand the basic principles of simulation, its applications, and the distinction between discrete and continuous systems.
CLO2	Build conceptual, specification, and computational models to represent real-world systems.
CLO3	Analyze simulation results, interpret statistical outputs, and evaluate performance metrics.
CLO4	Design and implement discrete-event simulation models using specialized tools and techniques.
CLO5	Assess system performance through simulation, identifying bottlenecks and exploring "what-if" scenarios.
CLO6	Validate and verify simulation models to ensure they accurately represent real-world systems.
CLO7	Use random variables, distributions, and probability functions to model uncertainties in simulations.
CLO8	Apply simulation techniques to solve practical problems in areas like manufacturing, logistics, and inventory management.

SUMMARY OF COURSE CONTENT

Sl.	Course Content	HRs	CLOs
1	Introduction to Simulation and System Modeling	3	CLO1, CLO2
2	Advantages and Disadvantages of Simulation	2	CLO1, CLO5
3	Types of Simulations (Monte Carlo, Time-Stepped, Discrete-Event)	4	CLO1, CLO4
4	Monte Carlo Simulation: Concepts and Applications	3	CLO4, CLO7
5	Simulation Model Taxonomy (Static, Dynamic, Continuous, Discrete)	2	CLO2, CLO5
6	Discrete-Event Simulation (DES) Components and Examples	4	CLO4, CLO6
7	Statistical Aspects: Random Variables & Distributions	3	CLO3, CLO7
8	Chi-Square Test for Validation	2	CLO6, CLO7
9	Verification and Validation of Simulation Models	3	CLO6
10	Applications of Simulation in Real-World Problems	3	CLO5, CLO8
11	Case Studies: Queueing Systems and Inventory Models	2	CLO4, CLO5

- **Recommended Books:**

1. **Simulation Modeling and Analysis**: Author: Averill M. Law (5th Edition or latest).
2. **Discrete-Event System Simulation**: Authors: Jerry Banks, John S. Carson II, Barry L. Nelson, David M. Nicol.
3. **Theory of Modeling and Simulation**: Authors: Bernard P. Zeigler, Herbert Praehofer, Tag Gon Kim.

ASSESSMENT PATTERN

CIE- Continuous Internal Evaluation (90 Marks)

Bloom's Category Marks (out of 90)	Tests (45)	Assignments (15)	Quizzes (15)	Attendance (15)
Remember	5	03		
Understand	5	04	05	
Apply	15	05	05	
Analyze	10			
Evaluate	5	03	05	
Create	5			

SEE- Semester End Examination (60 Marks)

Bloom's Category	Test
Remember	7
Understand	7
Apply	20
Analyze	15
Evaluate	6
Create	5

COURSE PLAN

Week No	Topics	Teaching Learning Strategy(s)	Assessment Strategy(s)	Alignment to CLO
1	Introduction to Simulation and System Modeling	Lecture, Discussion, Real-World Examples	Quiz on simulation basics	CLO1, CLO2
2	When Simulation is Appropriate/Not Appropriate	Lecture, Group Discussion	Short written assignment	CLO1
3	Advantages and Disadvantages of Simulation	Lecture, Case Study Analysis	Group presentation	CLO1, CLO5
4	Types of Simulations: Overview	Lecture, Class Exercise (Simulation Examples)	Class quiz	CLO1, CLO4
5	Monte Carlo Simulation: Concept and Applications	Lecture, Numerical Examples, Problem Solving	Problem-solving assignment	CLO4, CLO7
6	Monte Carlo Simulation: Hands-On Practice	Practical Demonstration, Software Simulation	Practical evaluation	CLO4, CLO7
7	Simulation Model Taxonomy	Lecture, Diagram Practice	Diagram-based quiz	CLO2, CLO5
8	Discrete-Event Simulation (DES): Components	Lecture, Flowchart Design, Examples	Flowchart-based exercise	CLO4, CLO6
9	DES: Single-Server Queue Example	Simulation Software Lab, Group Activity	Practical evaluation	CLO4, CLO6
10	Statistical Aspects: Random Variables & Distributions	Lecture, Numerical Examples	Class test on distributions	CLO3, CLO7
11	Chi-Square Test for Validation	Lecture, Hands-On Calculation, Examples	Assignment on Chi-Square Test	CLO6, CLO7
12	Verification and Validation of Simulation Models	Lecture, Validation Exercises, Examples	Model validation report	CLO6
13	Performance Analysis in Simulation	Lecture, Case Study Discussion	Short report on performance metrics	CLO5
14	Applications of Simulation in Real-World Problems	Lecture, Case Study Presentation	Group presentation	CLO5, CLO8
15	Case Study: Queueing Systems and Inventory Models	Simulation Lab, Practical Exercises	Practical evaluation	CLO4, CLO5
16	Common Mistakes in Simulation	Lecture, Group Activity	Short reflective report	CLO3, CLO6
17	Review and Final Assessment	Revision, Q&A Session	Final Exam	CLO1, CLO2, CLO3, CLO4, CLO5, CLO6, CLO7, CLO8

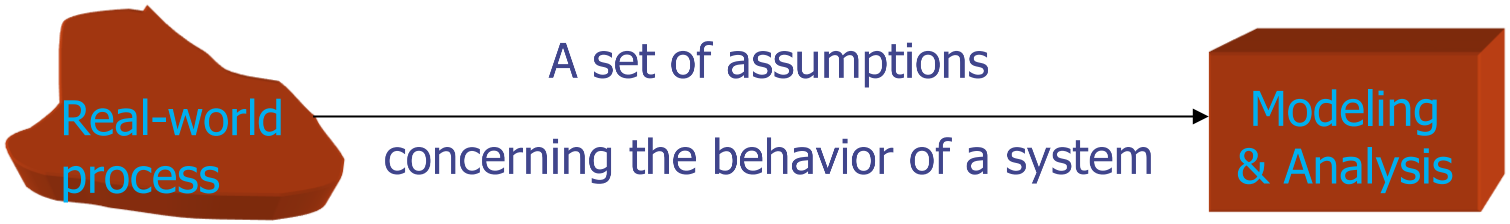
WEEK 1
SLIDES 6-13

LECTURE 1

CHAPTER-01
DISCRETE-EVENT SYSTEM SIMULATION

Introduction to Simulation

INTRODUCTION TO SIMULATION



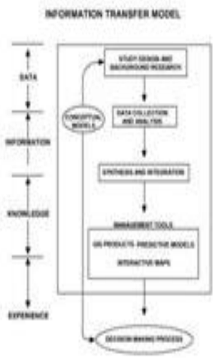
- Simulation
  - the imitation of the operation of a real-world process or system over time
  - to develop a set of assumptions of mathematical, logical, and symbolic relationship between the entities of interest, of the system.
  - to estimate the measures of performance of the system with the simulation-generated data

- Simulation modeling can be used
  - as an analysis tool for predicting the effect of changes to existing systems
  - as a design tool to predict the performance of new systems

Miracle on the Hudson in 2009 with 155 passengers

- A computer simulation is an application designed to imitate a real-life situation. A good example is software which simulates the experience of piloting a plane.


200 m
42 km/h
D9
Wed 8:19
Navigation
€32,439.-
F5
ETA: Wed 13:01, 291 km, 4 h 42 min
F1
F5
F6
F7
F8
Route Advisor



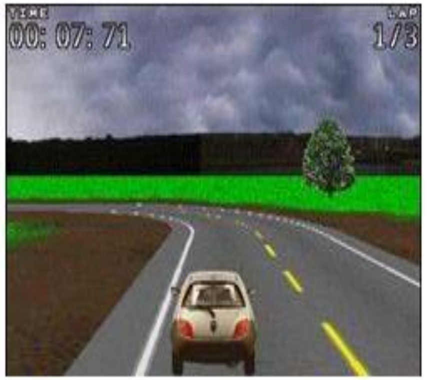
The advantages of **modelling**

Much cheaper than using the real thing.

Quick results. Simulation can be speeded up.

Can simulate dangerous situations.

Allows for data to be changed to see the consequences.



WEEK 2
SLIDES 14-23

WHEN SIMULATION IS THE APPROPRIATE TOOL (1)

- Simulation enables the study of, and experimentation with, the internal interactions of a complex system, or of a subsystem within a complex system.
- Informational, organizational, and environmental changes can be simulated, and the effect of these alterations on the model's behavior can be observed.
- The knowledge gained in designing a simulation model may be of great value toward suggesting improvement in the system under investigation.
- By changing simulation inputs and observing the resulting outputs, valuable insight may be obtained into which variables are most important and how variables interact.
- Simulation can be used as a pedagogical device to reinforce analytic solution methodologies.

WHEN SIMULATION IS THE APPROPRIATE TOOL (2)

- Simulation can be used to experiment with new designs or policies prior to implementation, so as to prepare for what may happen.
- Simulation can be used to verify analytic solutions.
- By simulating different capabilities for a machine, requirements can be determined.
- Simulation models designed for training allow learning without the cost and disruption of on-the-job learning.
- Animation shows a system in simulated operation so that the plan can be visualized.
- The modern system (factory, wafer fabrication plant, service organization, etc.) is so complex that the interactions can be treated only through simulation.

WHEN SIMULATION IS NOT APPROPRIATE

- When the problem can be solved using common sense.
- When the problem can be solved analytically.
- When it is easier to perform direct experiments.
- When the simulation costs exceed the savings.
- When the resources or time are not available.
- When system behavior is too complex or can’t be defined.
- When there isn’t the ability to verify and validate the model.

ADVANTAGES AND DISADVANTAGES OF SIMULATION (1)

- Advantages
  - New polices, operating procedures, decision rules, information flows, organizational procedures, and so on can be explored without disrupting ongoing operations of the real system.
  - New hardware designs, physical layouts, transportation systems, and so on, can be tested without committing resources for their acquisition.
  - Hypotheses about how or why certain phenomena occur can be tested for feasibility.
  - Insight can be obtained about the interaction of variables.
  - Insight can be obtained about the importance of variables to the performance of the system.
  - Bottleneck analysis can be performed indicating where work-in-process, information, materials, and so on are being excessively delayed.
  - A simulation study can help in understanding how the system operates rather than how individuals think the system operates.
  - “What-if” questions can be answered. This is particularly useful in the design of new system.

ADVANTAGES AND DISADVANTAGES OF SIMULATION (2)

- Disadvantages
  - Model building requires special training. It is an art that is learned over time and through experience. Furthermore, if two models are constructed by two competent individuals, they may have similarities, but it is highly unlikely that they will be the same.
  - Simulation results may be difficult to interpret. Since most simulation outputs are essentially random variables (they are usually based on random inputs), it may be hard to determine whether an observation is a result of system interrelationships or randomness.
  - Simulation modeling and analysis can be time consuming and expensive. Skimping on resources for modeling and analysis may result in a simulation model or analysis that is not sufficient for the task.
  - Simulation is used in some cases when an analytical solution is possible, or even preferable. This might be particularly true in the simulation of some waiting lines where closed-form queueing models are available.

AREAS OF APPLICATION (1)

- WSC(Winter Simulation Conference) : http://www.wintersim.org
  - Manufacturing Applications
    - Analysis of electronics assembly operations
    - Design and evaluation of a selective assembly station for high-precision scroll compressor shells
    - Comparison of dispatching rules for semiconductor manufacturing using large-facility models
    - Evaluation of cluster tool throughput for thin-film head production
    - Determining optimal lot size for a semiconductor back-end factory
    - Optimization of cycle time and utilization in semiconductor test manufacturing
    - Analysis of storage and retrieval strategies in a warehouse
    - Investigation of dynamics in a service-oriented supply chain
    - Model for an Army chemical munitions disposal facility
  - Semiconductor Manufacturing
    - Comparison of dispatching rules using large-facility models
    - The corrupting influence of variability
    - A new lot-release rule for wafer fabs

Areas of Application (2)

  - Assessment of potential gains in productivity due to proactive reticle management
  - Comparison of a 200-mm and 300-mm X-ray lithography cell
  - Capacity planning with time constraints between operations
  - 300-mm logistic system risk reduction
- **Construction Engineering**
  - Construction of a dam embankment
  - Trenchless renewal of underground urban infrastructures
  - Activity scheduling in a dynamic, multi-project setting
  - Investigation of the structural steel erection process
  - Special-purpose template for utility tunnel construction
- **Military Application**
  - Modeling leadership effects and recruit type in an Army recruiting station
  - Design and test of an intelligent controller for autonomous underwater vehicles
  - Modeling military requirements for nonwarfighting operations
  - Multi-trajectory performance for varying scenario sizes
  - Using adaptive agent in U.S Air Force pilot retention

Areas of Application (3)

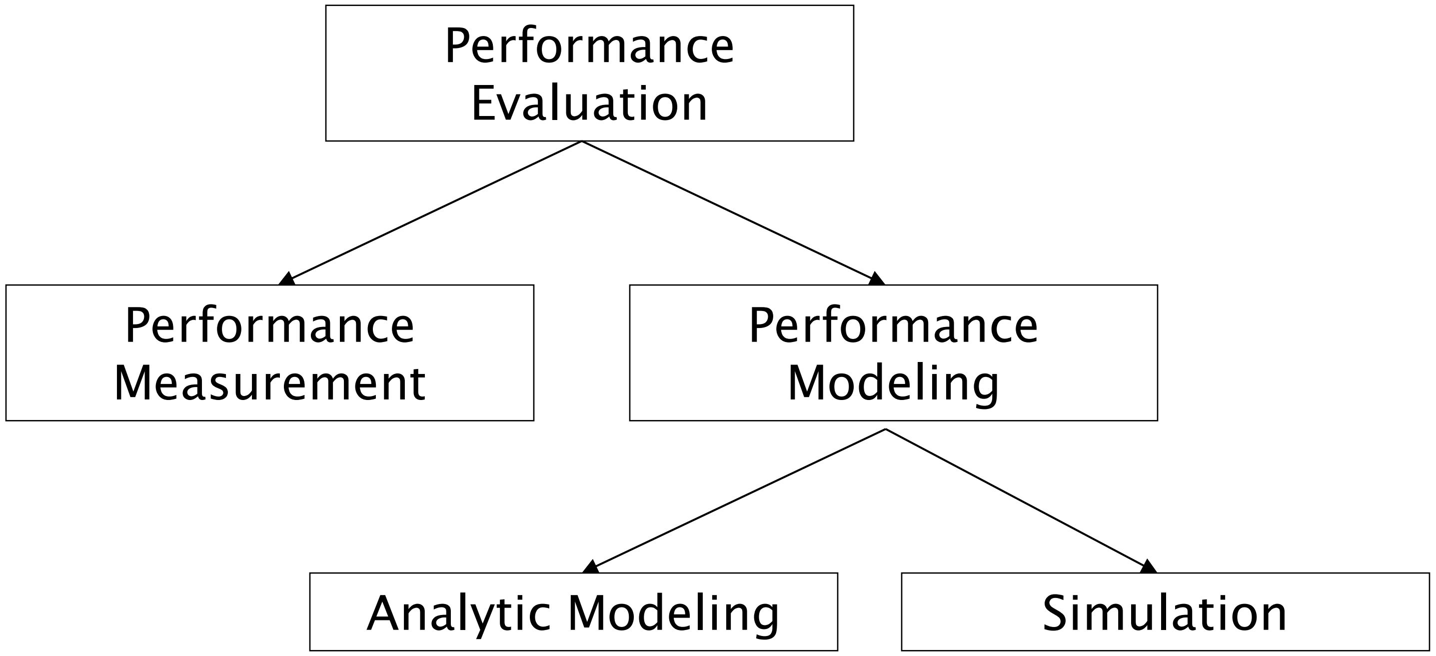
- Logistics, Transportation, and Distribution Applications
  - Evaluating the potential benefits of a rail-traffic planning algorithm
  - Evaluating strategies to improve railroad performance
  - Parametric modeling in rail-capacity planning
  - Analysis of passenger flows in an airport terminal
  - Proactive flight-schedule evaluation
  - Logistics issues in autonomous food production systems for extended-duration space exploration
  - Sizing industrial rail-car fleets
  - Product distribution in the newspaper industry
  - Design of a toll plaza
  - Choosing between rental-car locations
  - Quick-response replenishment

Areas of Application (4)

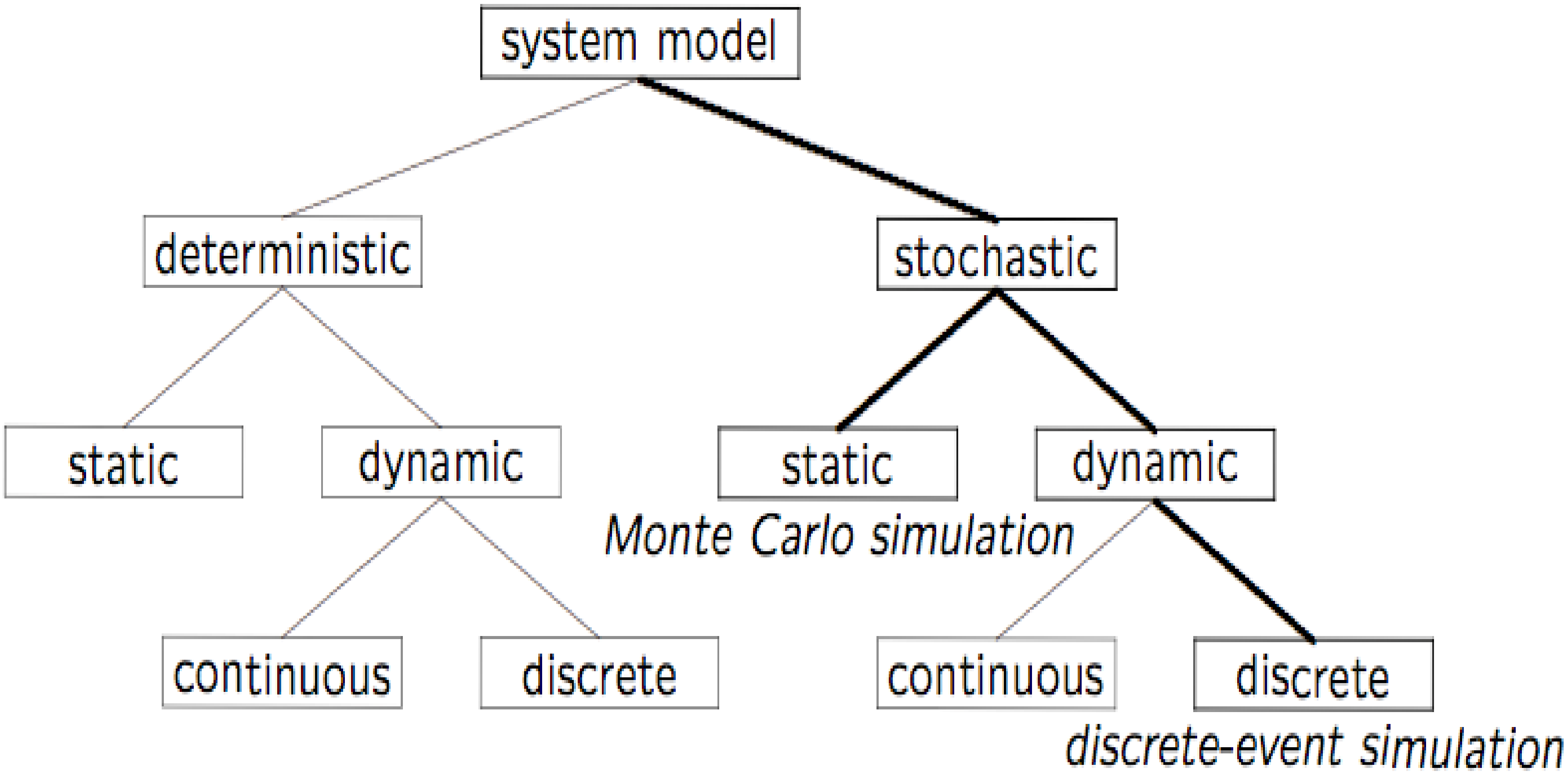
- Business Process Simulation
  - Impact of connection bank redesign on airport gate assignment
  - Product development program planning
  - Reconciliation of business and systems modeling
  - Personnel forecasting and strategic workforce planning
- Human Systems
  - Modeling human performance in complex systems
  - Studying the human element in air traffic control

WEEK 3
SLIDES 24-31

RECAP: PERFORMANCE EVALUATION



SIMULATION MODEL TAXONOMY (PREVIEW)



TERMINOLOGY (1 OF 2)

- A *system* is defined as a group of objects that interact with each other to accomplish some purpose
  - A computer system: CPU, memory, disk, bus, NIC
  - An automobile factory: Machines, components parts and workers operate jointly along assembly line
- A system is often affected by changes occurring outside the system: system environment
  - Hair salon: arrival of customers
  - Warehouse: arrival of shipments, fulfilling of orders
    - Effect of supply on demand: relationship between factory output from supplier and consumption by customers

TERMINOLOGY (2 OF 2)

- Entity
  - An object of interest in the system: Machines in factory
- Attribute
  - The property of an entity: speed, capacity, failure rate
- State
  - A collection of variables that describe the system in any time: status of machine (busy, idle, down,…)
- Event
  - An instantaneous occurrence that might change the state of the system: breakdown

SIMULATION MODELING

- Develop a simulation program that implements a computational model of the system of interest
- Run the simulation program and use the data collected to estimate the performance measures of interest (often involves the use of randomization)
- A system can be studied at an arbitrary level of detail

- Quote of the day:

“The hardest part about simulation is deciding what NOT to model.”
- Moe Lavigne, Stentor, Summer 1995

ADVANTAGES OF SIMULATION

- New policies and procedures can be explored without disrupting the ongoing operation of the real system
- New designs can be tested without committing resources for their acquisition
- Time can be compressed or expanded to allow for a speed-up or slow-down of the phenomenon under study
- Insight can be obtained about the interactions of variables, and which ones have the most impact on system performance
- Can obtain answers to "What if…" questions

DISADVANTAGES OF SIMULATION

- Model building requires special training
  - An important role for courses like CPSC 531!!
  - Vendors of simulation software have been actively developing packages that contain models that only need input (templates), which simplifies things for users
- Simulation results can be difficult to interpret
  - Need proper statistical interpretation for output analysis
- Simulation modeling and analysis can be time- consuming and expensive, both for the modeler, as well as in compute time (if not done judiciously)

WEEK 4
SLIDES 32-41

WHEN SIMULATION IS NOT APPROPRIATE

- When the problem can be solved by common sense
- When the problem can be solved analytically
- When it is easier to perform direct experiments
- When cost of simulations exceeds (expected) savings for the real system
- When system behavior is too complex (e.g., humans)

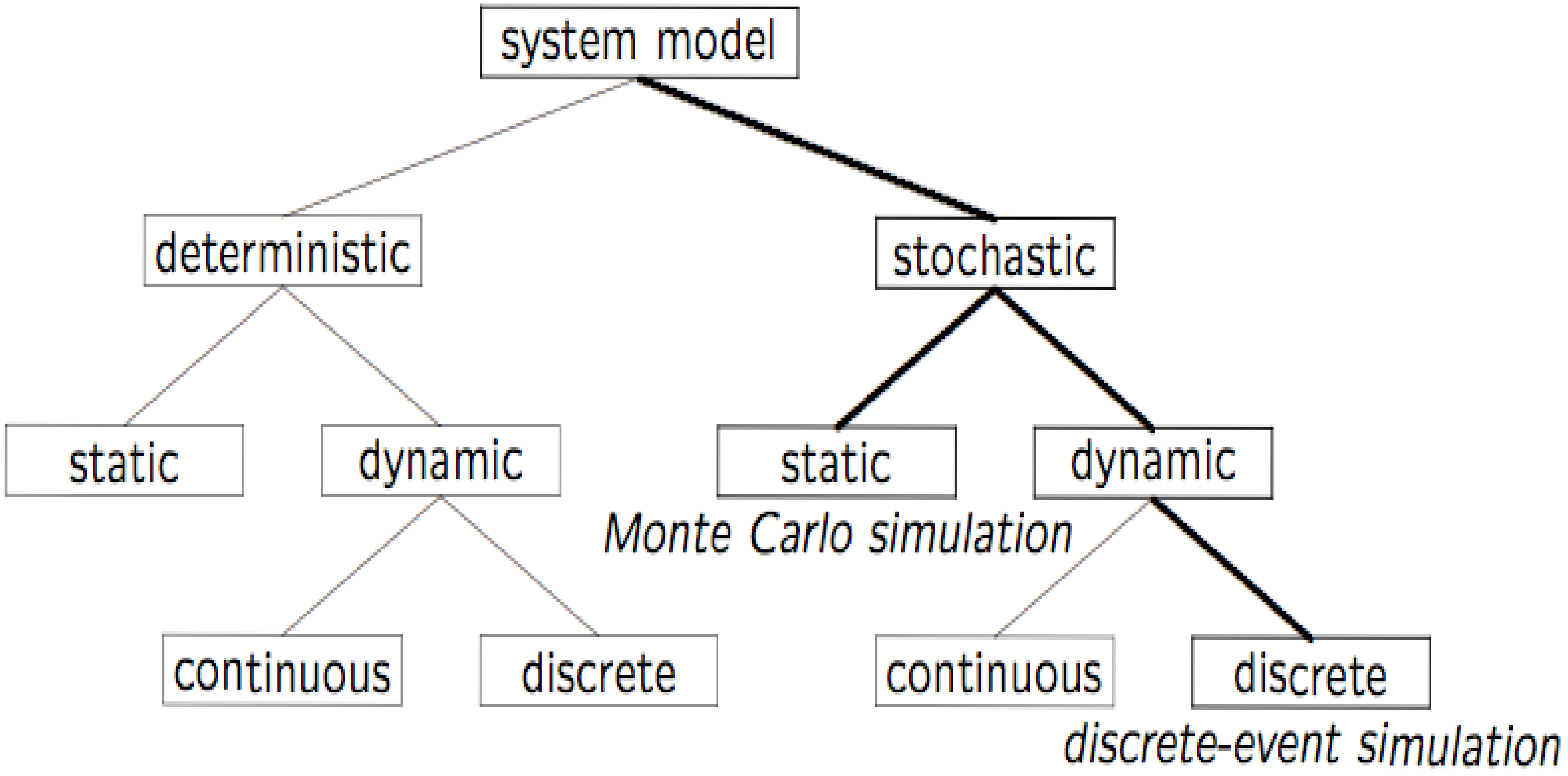
COMMON MISTAKES IN SIMULATION

- Poor (pseudo) random number generators
  - Best to use well-known or well-understood generator
- Improper selection of seeds for PRNG
  - Short periods; same seeds for all streams
- Inappropriate level of detail:
  - More detail → more time → more bugs
  - More parameters $\neq$ more accurate
- Improperly handled initial conditions (warmup)
- Improperly handled ending conditions (cooldown)
- Run-length too short to achieve steady-state
  - Need proper output analysis, confidence intervals

TYPES OF SIMULATIONS

- Monte Carlo simulation
- Time-stepped simulation
- Trace-driven simulation
- Discrete-event simulation
- Continuous simulation

SIMULATION MODEL TAXONOMY



SIMULATION EXAMPLES

- Monte Carlo simulation
  - Estimating π
  - Craps (dice game)
- Time-stepped simulation
  - Mortgage scenarios
- Trace-driven simulation
  - Single-server queue (ssq1.c)
- Discrete-event simulation
  - Witchcraft hair salon

MONTE CARLO SIMULATION

Named after Count Montgomery de Carlo, who was a famous Italian gambler and random-number generator (1792-1838).

- Static simulation (no time dependency)
- To model probabilistic phenomenon
- Can be used for evaluating non-probabilistic expressions using probabilistic methods
- Can be used for estimating quantities that are "hard" to determine analytically or experimentally

TRACE-DRIVEN SIMULATION

- Trace = time-ordered record of events in system
- Trace-driven simulation = Trace input
- Often used in evaluating or tuning resource management algorithms (based on real workloads):
  - Paging, cache analysis, CPU scheduling, deadlock prevention, dynamic storage allocation
- Example: Trace = start time + duration of processes
- Example: Trace = size in bytes of file written to disk
- Example: Trace = mobile device ID and call duration

ADVANTAGES OF TRACE-DRIVEN SIMULATIONS

- Credibility
- Easy validation: compare simulation with measurement
- Accurate workload: models correlation and interference
- Fair comparison: better than random input
- Similarity to the actual implementation:
  - trace-driven model is similar to the system
  - can understand complexity of implementation

DISADVANTAGES OF TRACE-DRIVEN SIMULATIONS

- Complexity: more detailed
- Representativeness: workload changes with time, equipment
- Data Collection: few minutes fill up a disk
- Instrumentation: granularity; intrusiveness
- Single Point of Validation: one trace = one point
- Difficult to change workload

WEEK 5

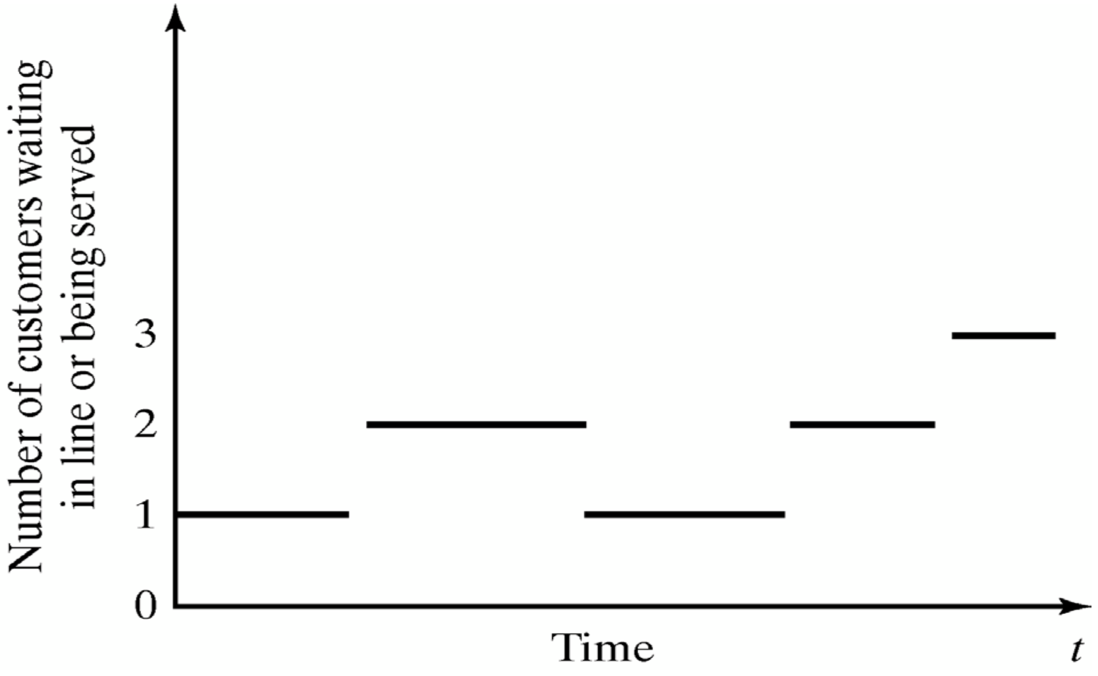
SLIDES 42-49

DISCRETE-EVENT SIMULATION

- A simulation model with three features:
  1. Stochastic:
     some variables in the simulation model are random
  2. Dynamic:
     system state evolves over time
  3. Discrete-Event:
     changes in system state occur at discrete time instances

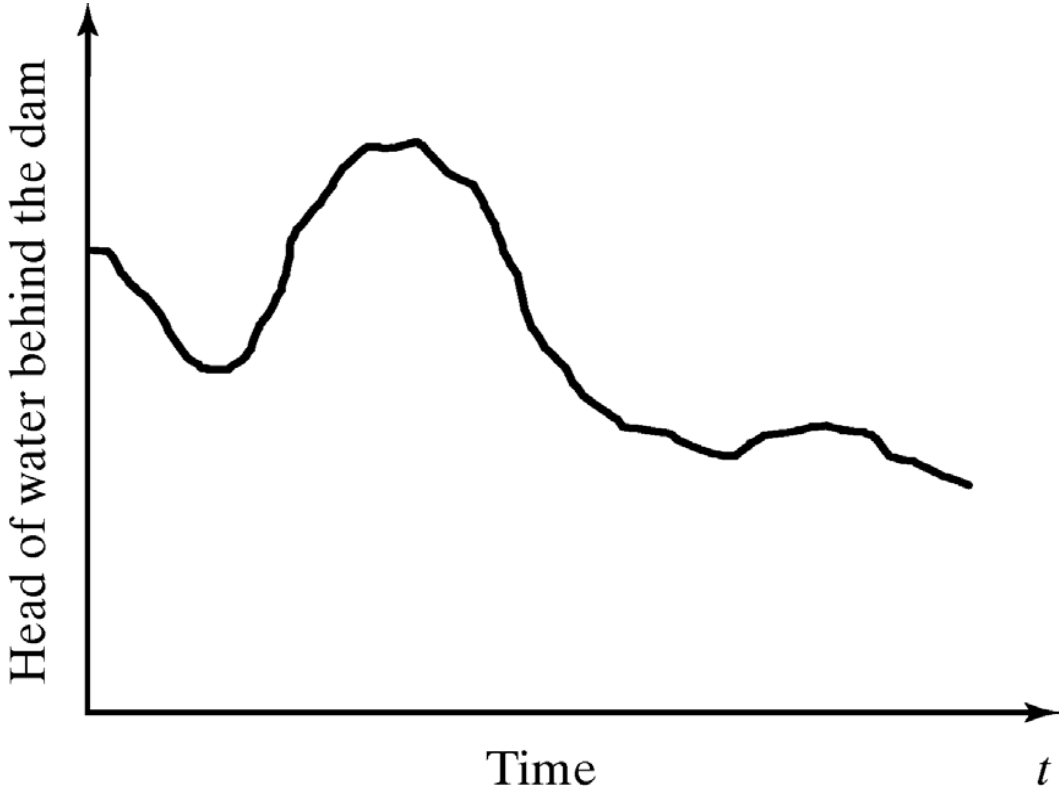
DISCRETE AND CONTINUOUS SYSTEMS

- A discrete system is one in which the system state changes only at a discrete set of points in time
  - Example: A restaurant



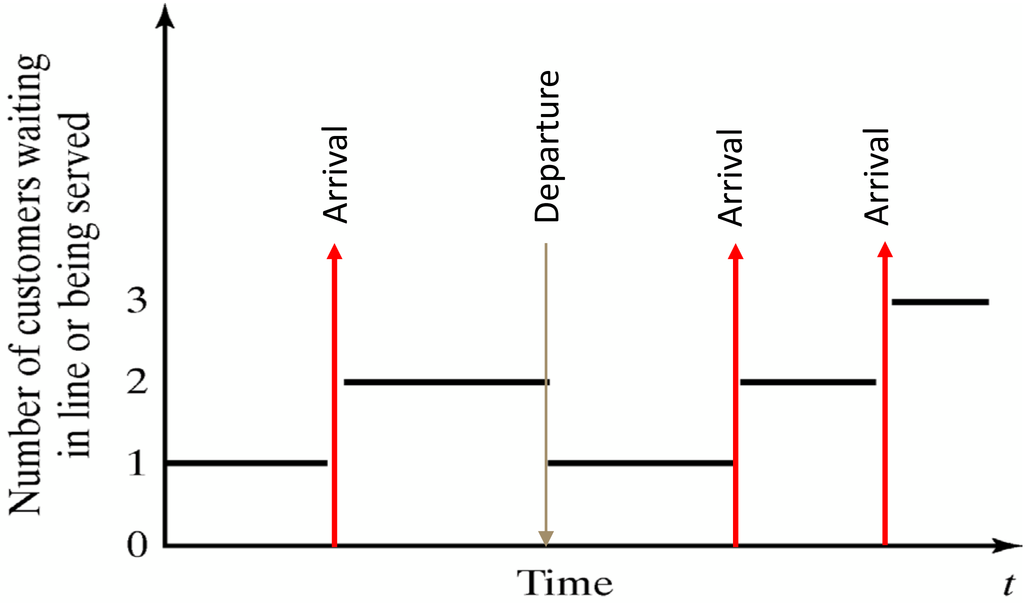
DISCRETE AND CONTINUOUS SYSTEMS

- A continuous system is one in which the system state changes continuously over time
  - Example: Water level in Bow River (or Bearspaw dam)



DISCRETE-EVENT SIMULATION

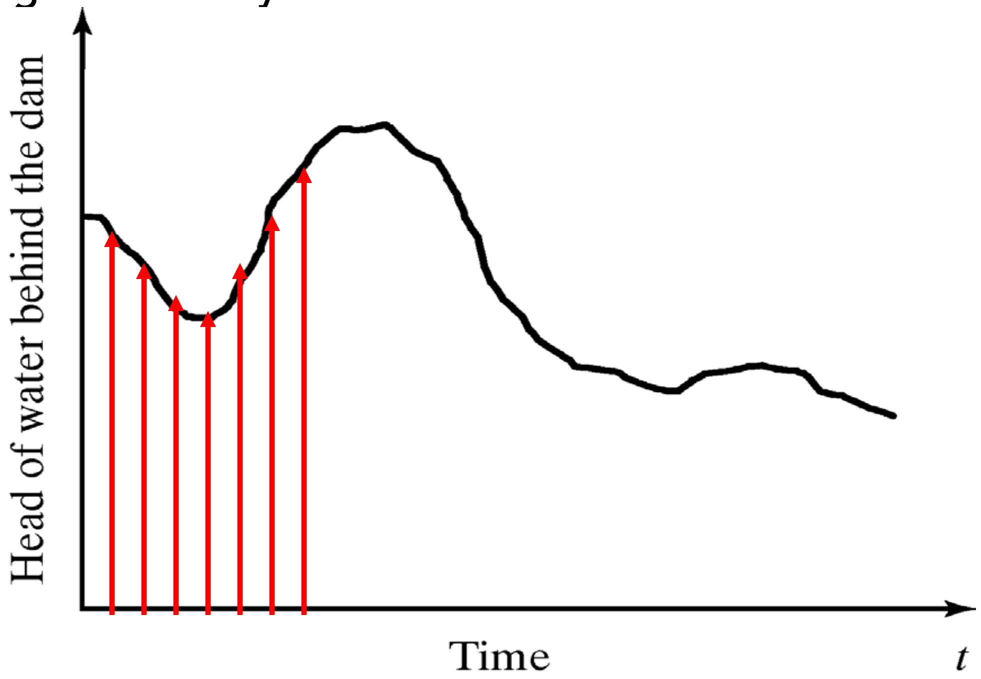
- A simulation model in which system state evolves over a discrete sequence of events in time
  - System state changes only when an event occurs
  - System state does not change between the events



Restaurant Example

CONTINUOUS SIMULATION

- A simulation model in which system state evolves continuously over time
  - Time is divided to small time slices
  - System state changes in every time slice

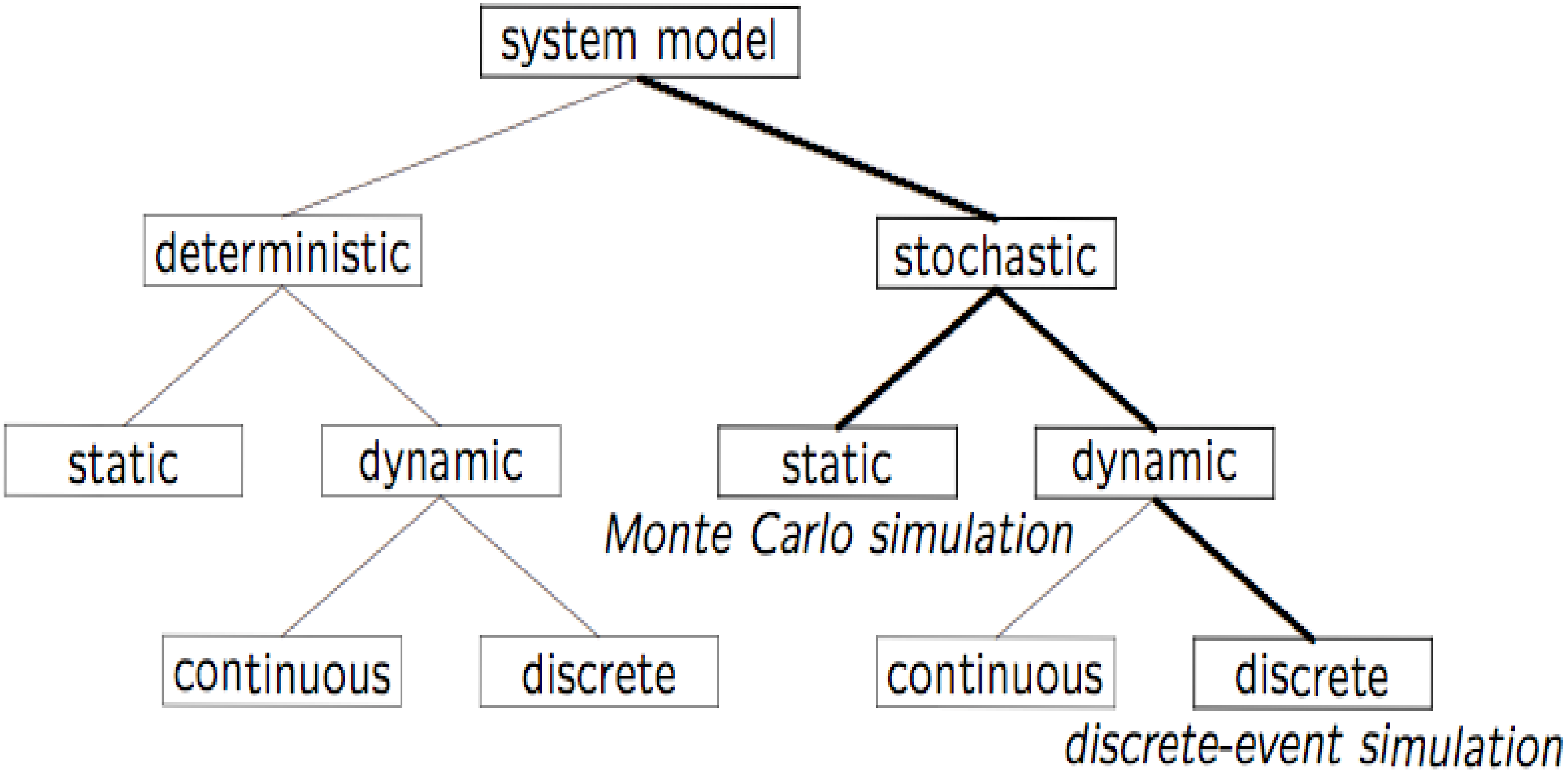


Dam Example

CHARACTERIZING A SIMULATION MODEL

- Deterministic or Stochastic
  - Does the model contain stochastic components?
- Static or Dynamic
  - Is time a significant variable?
- Continuous or Discrete
  - Does the system state evolve continuously or only at discrete points in time?

SIMULATION MODEL TAXONOMY



WEEK 6
SLIDES 50-54

DES MODEL DEVELOPMENT

- How to develop a simulation model:
  1. Determine the goals and objectives
  2. Build a ***conceptual*** model
  3. Convert into a ***specification*** model
  4. Convert into a ***computational*** model
  5. Verify the model
  6. Validate the model
- Typically an iterative process

THREE MODEL LEVELS

- Conceptual Model
  - Very high level (perhaps schematic diagram)
  - How comprehensive should the model be?
  - What are the state variables?
  - Which ones are dynamic, and which are most important?
- Specification Model
  - On paper: entitites, interactions, requirements, rules, etc.
  - May involve equations, pseudocode, etc.
  - How will the model receive input?
- Computational Model
  - A computer program
  - General-purpose programming language or simulation language?

SIMULATION SOFTWARE

- General purpose programming languages
  - Flexible and familiar
  - Well suited for learning DES principles and techniques
  - E.g., C++, Java
- Simulation programming languages
  - Good for building models quickly
  - Provide built-in features (e.g., queue structures)
  - Graphics and animation provided
  - Domain specific
    - Network protocol simulation: ns2, Opnet
    - Electrical power simulation: ETAP
    - Design and engineering: Ansys, Autodesk
    - Process simulation: Simul8

VERIFICATION AND VALIDATION

- Verification
  - Computational model should be consistent with specification model
  - Did we build the model right?
- Validation
  - Computational model should be consistent with the system being analyzed
  - Did we build the right model?
  - Can an expert distinguish simulation output from system output?

WEEK 7
SLIDES 55-60

COMPONENTS AND ORGANIZATION OF A DISCRETE-EVENT SIMULATION MODEL

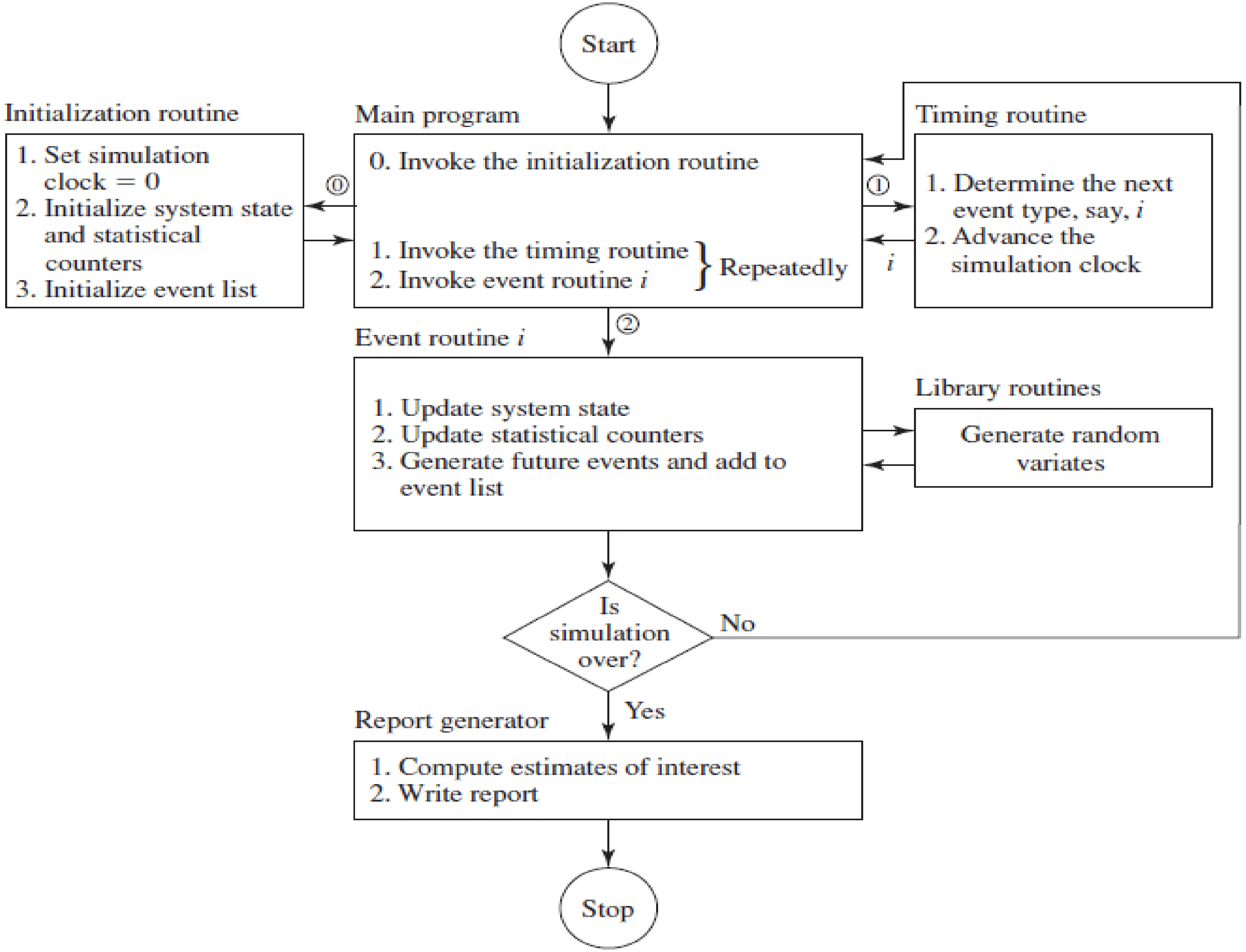
COMPONENTS OF A DISCRETE EVENT SIMULATION MODEL

- System state
  - The collection of state variables necessary to describe the system at a particular time
- Simulation clock
  - A variable giving the current value of simulated time
- Event list
  - A list containing the next time when each type of event will occur
- Statistical counters
  - Variables used for storing statistical information about system performance

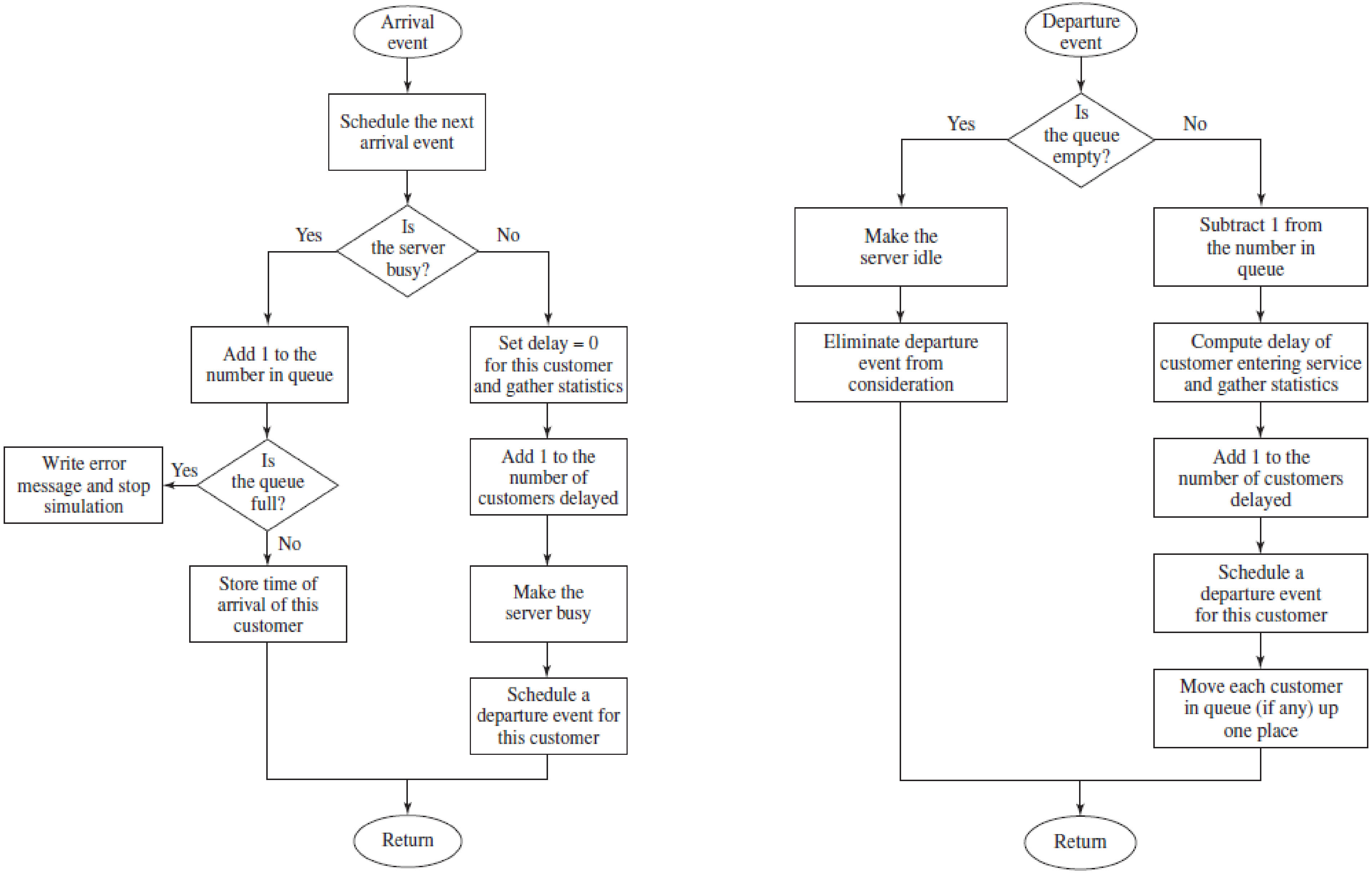
COMPONENTS OF A DISCRETE EVENT SIMULATION MODEL

- Initialization routine
  - A subprogram to initialize the simulation model at time 0
- Timing routine
  - A subprogram that determines the next event from the event list and then advances the simulation clock to the time when that event is to occur
- Event routine
  - A subprogram that updates the system state when a particular type of event occurs (there is one event routine for each event type)
- Library routines
  - A set of subprograms used to generate random observations from probability distributions that were determined as part of the simulation model

ORGANIZATION OF A DISCRETE EVENT SIMULATION MODEL



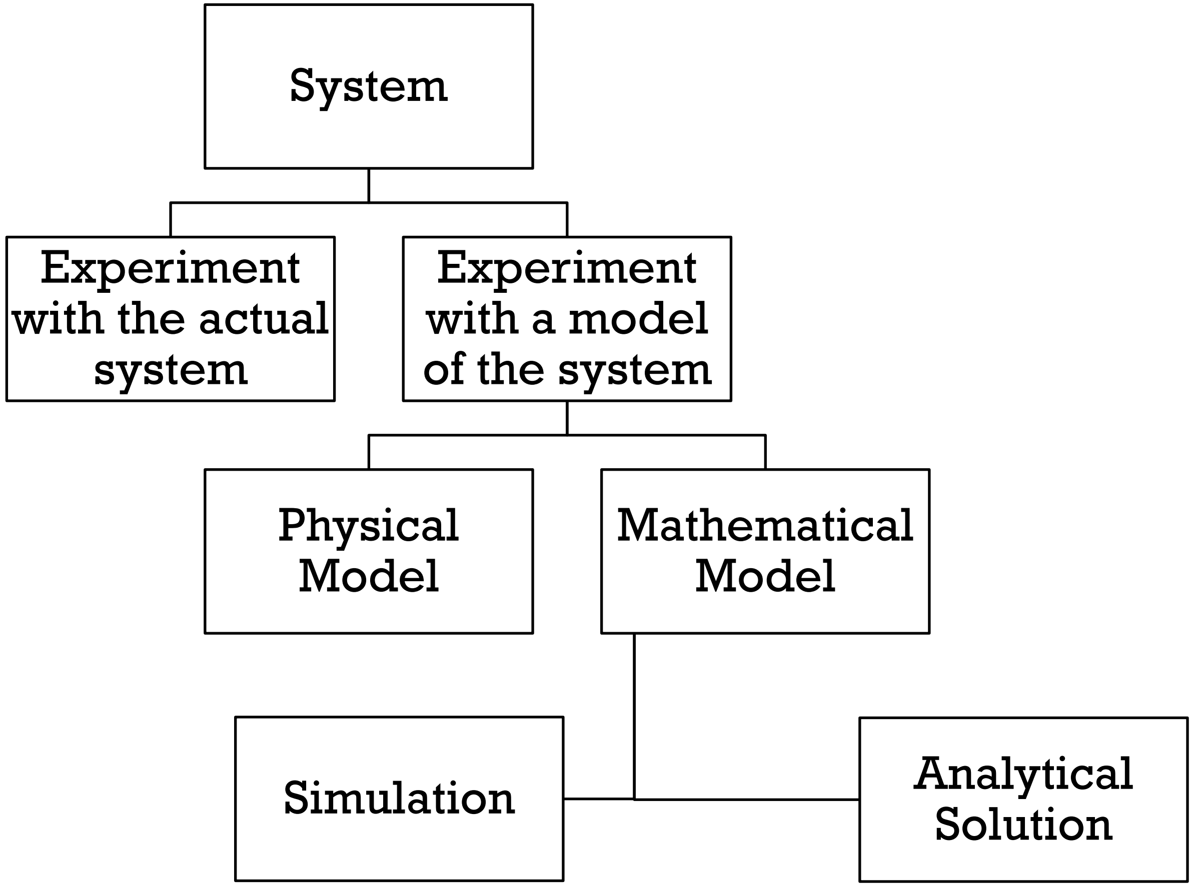
FLOWCHARTS FOR ARRIVAL AND DEPARTURE ROUTINES



WEEK 8
SLIDES 61-101

BASIC SIMULATION MODELING

WAYS TO STUDY A SYSTEM

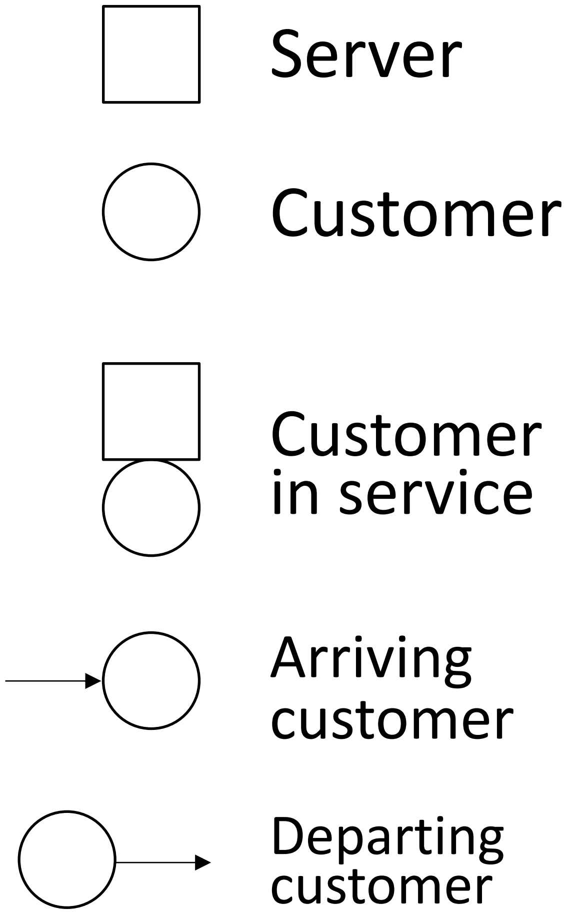


64

DISCRETE EVENT SIMULATION

Md. Masudur Rahman, Dept. of CSE, UGV

SINGLE SERVER QUEUING SYSTEM



- t_i = time of arrival of the ith customer ($t_0 = 0$)
- $A_i = t_i - t_{i-1}$ = interarrival time between $(i-1)$st and ith arrivals of customers
- S_i = time that server actually spends serving ith customer (exclusive of customer's delay in queue)
- D_i = delay in queue of ith customer (or waiting time)
- $c_i = t_i + D_i + S_i$ = time that ith customer completes service and departs
- e_i = time of occurrence of ith event of any type (ith value the simulation clock takes on, excluding the value $e_0 = 0$)

SINGLE SERVER QUEUING SYSTEM

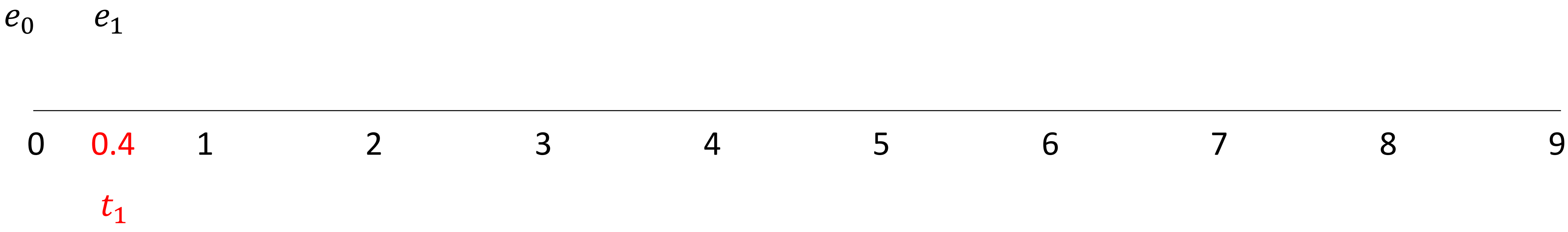
- Given
- t_i =0.4, 1.6, 2.1, 3.8, 4.0, 5.6, 5.8, and 7.2
- c_i =2.4, 3.1, 3.3, 4.9, and 8.6
- End time = 8.6

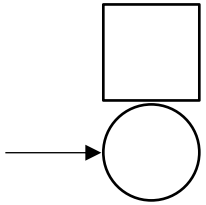
SIMULATION START...

e_0

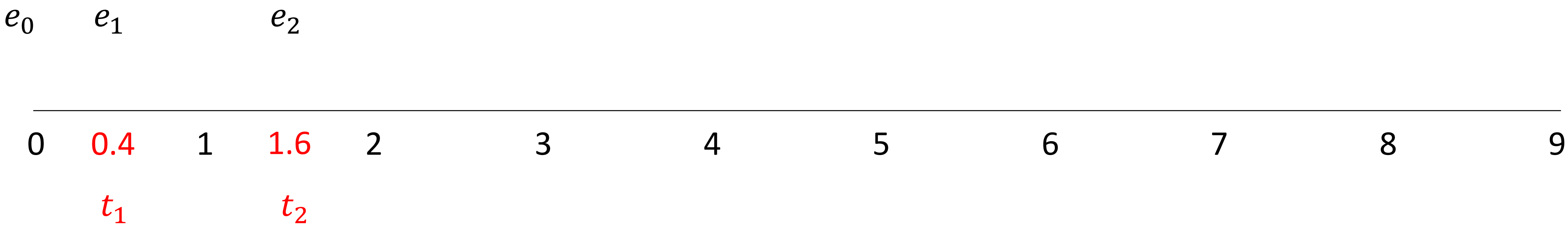
0 1 2 3 4 5 6 7 8 9

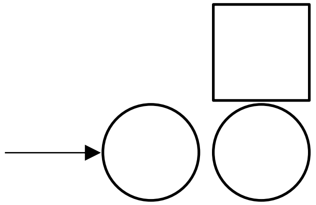
CUSTOMER ARRIVES, TAKES SERVICE…



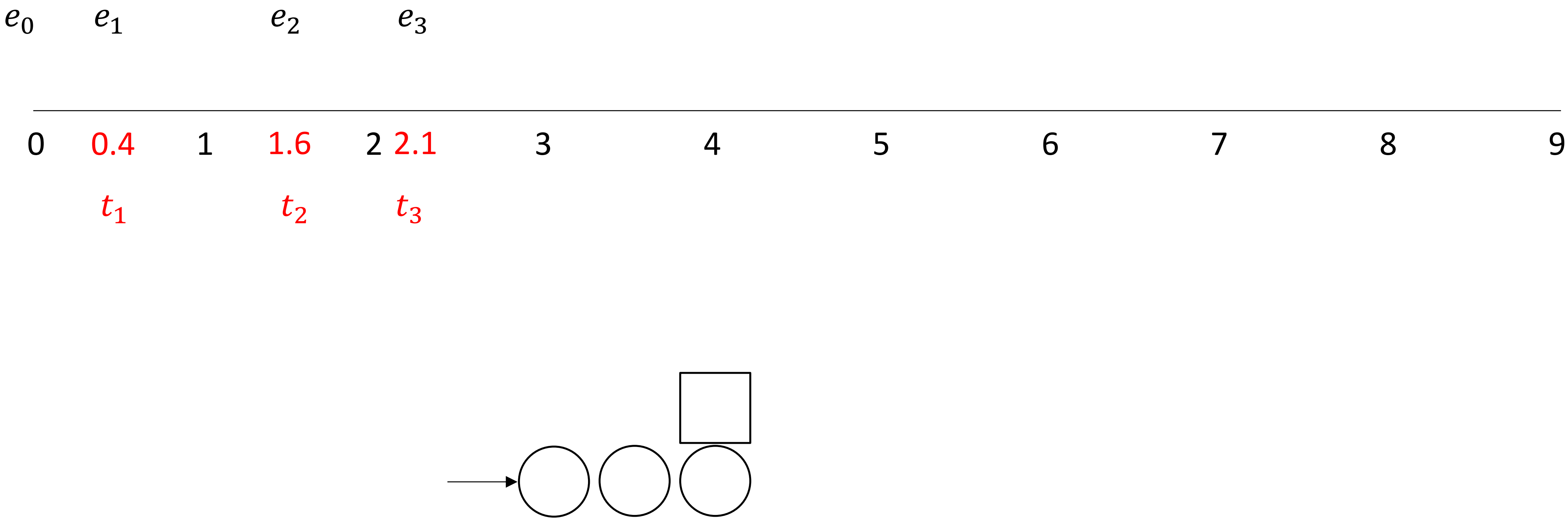


ANOTHER CUSTOMER ARRIVES AND WAITS…

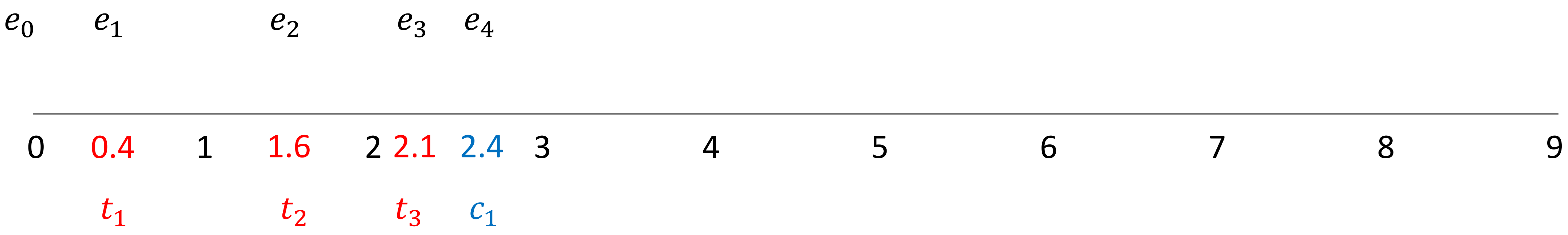


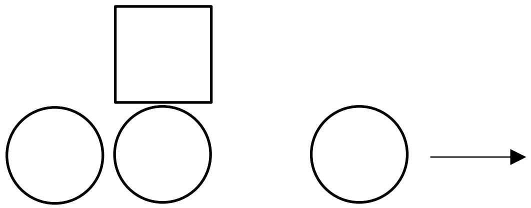


ANOTHER CUSTOMER ARRIVES AND WAITS...

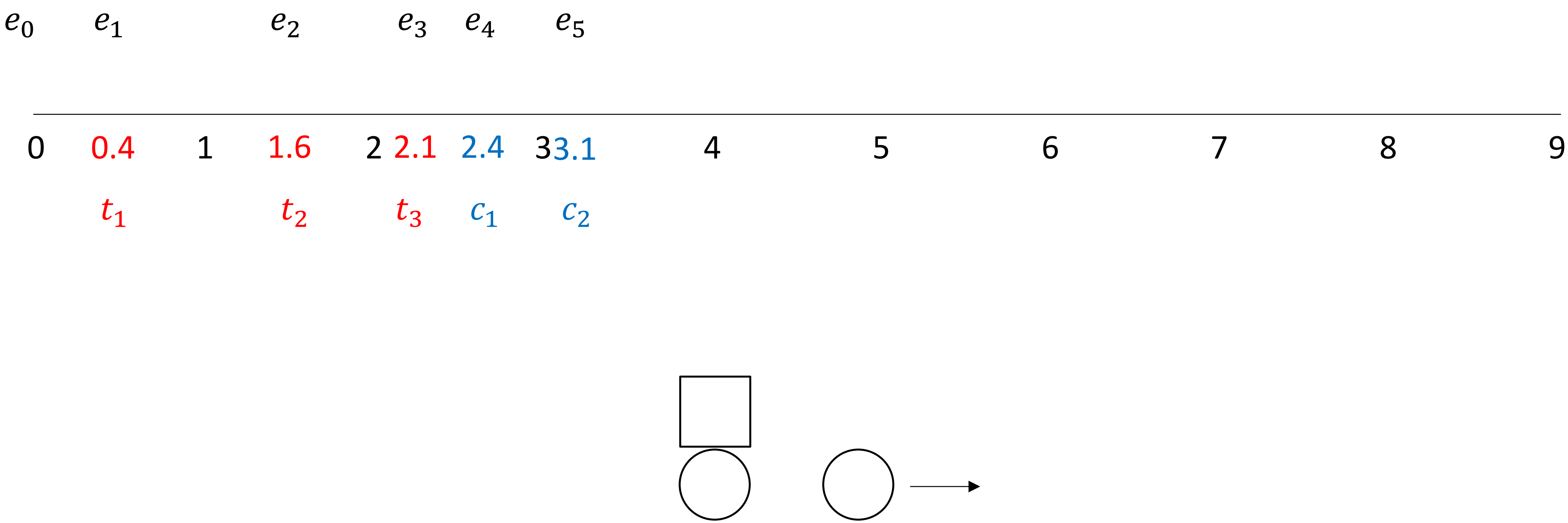


A CUSTOMER LEAVES, ANOTHER TAKES SERVICE...

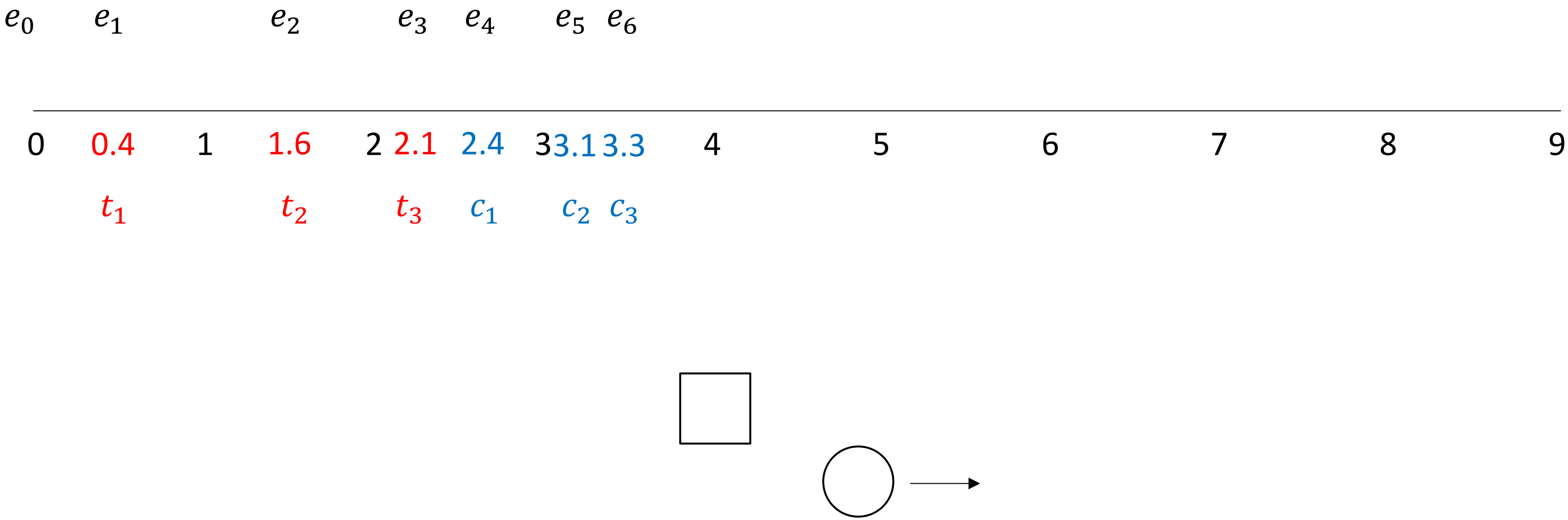




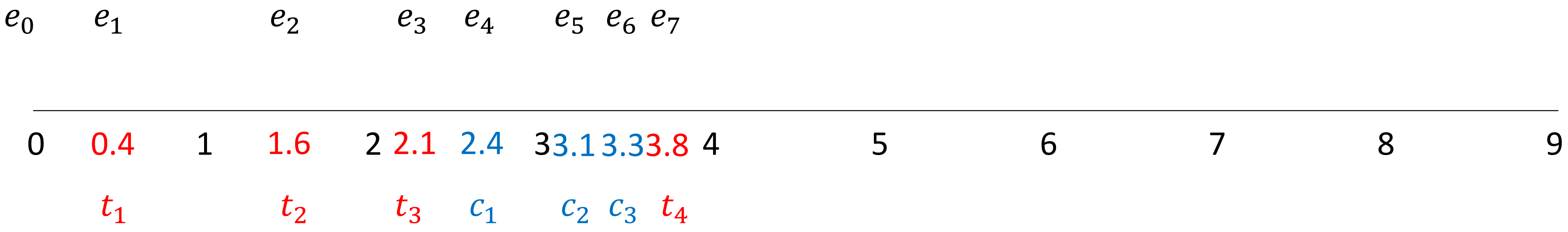
ANOTHER CUSTOMER LEAVES...

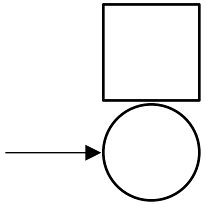


ANOTHER CUSTOMER LEAVES...



CUSTOMER ARRIVES…



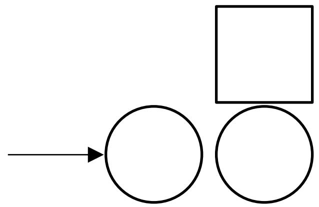


CUSTOMER ARRIVES...

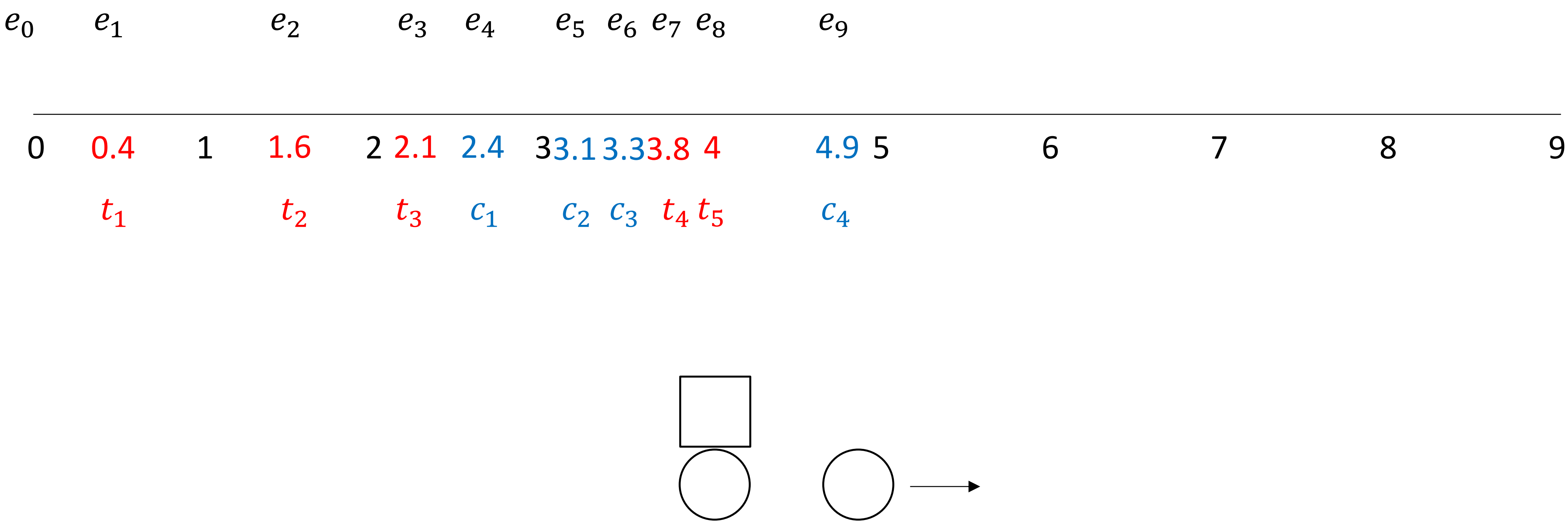
e_0 e_1 e_2 e_3 e_4 e_5 e_6 e_7 e_8

0 0.4 1 1.6 2 2.1 2.4 3 3.1 3.3 3.8 4 5 6 7 8 9

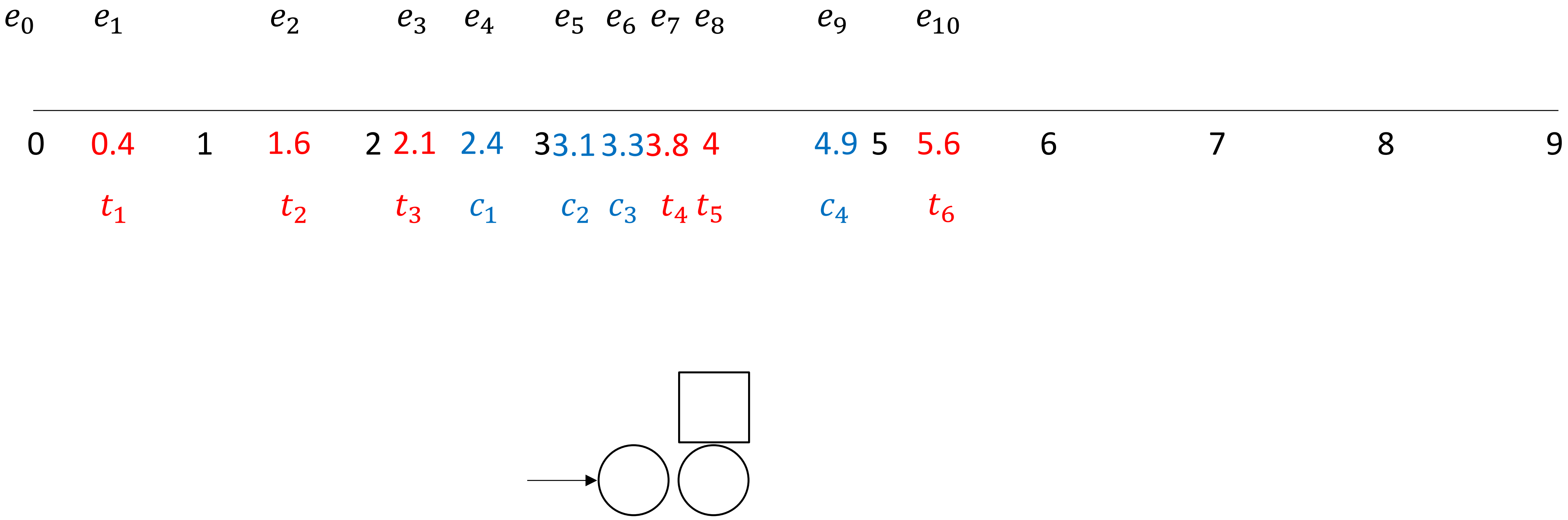
t_1 t_2 t_3 c_1 c_2 c_3 t_4 t_5



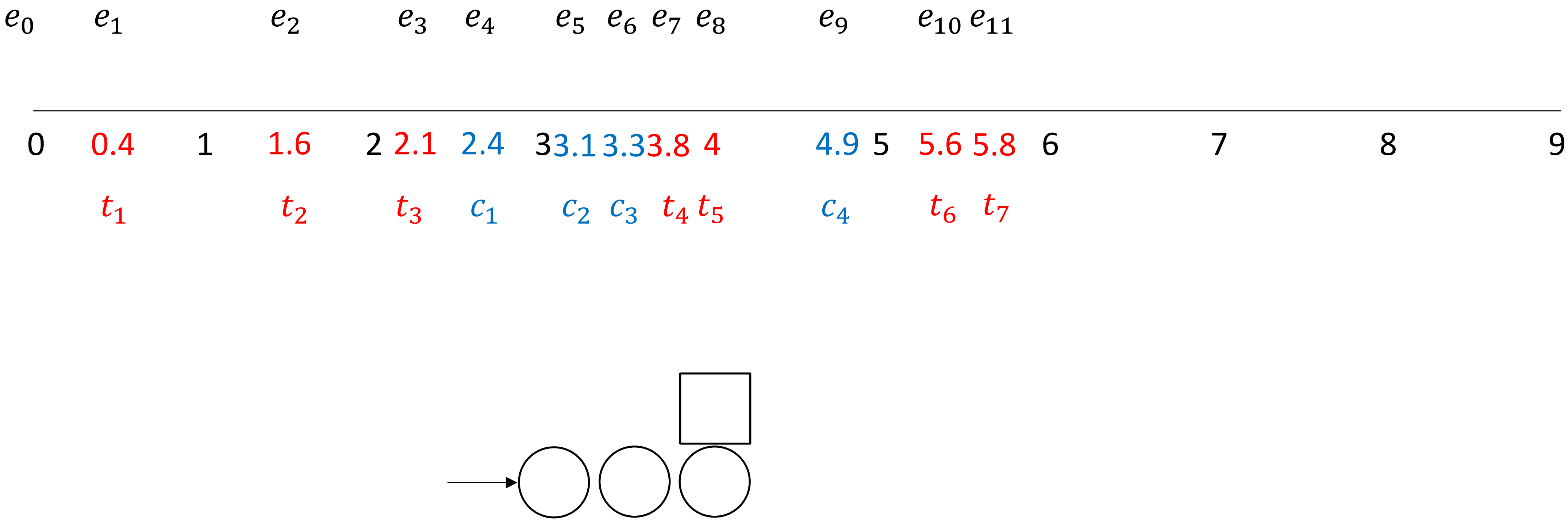
CUSTOMER LEAVES...



CUSTOMER ARRIVES...



CUSTOMER ARRIVES...

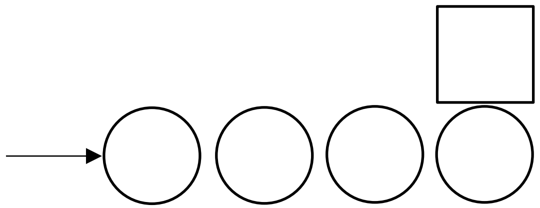


CUSTOMER ARRIVES...

e_0 e_1 e_2 e_3 e_4 e_5 e_6 e_7 e_8 e_9 e_{10} e_{11} e_{12}

0 0.4 1 1.6 2 2.1 2.4 3 3.1 3.3 3.8 4 4.9 5 5.6 5.8 6 7 7.2 8 9

t_1 t_2 t_3 c_1 c_2 c_3 t_4 t_5 c_4 t_6 t_7 t_8

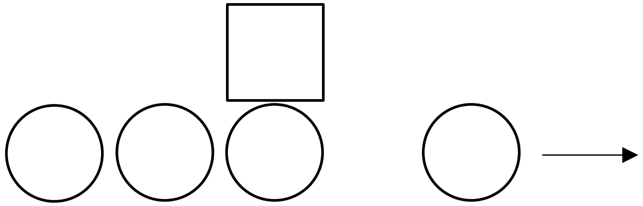


CUSTOMER LEAVES…

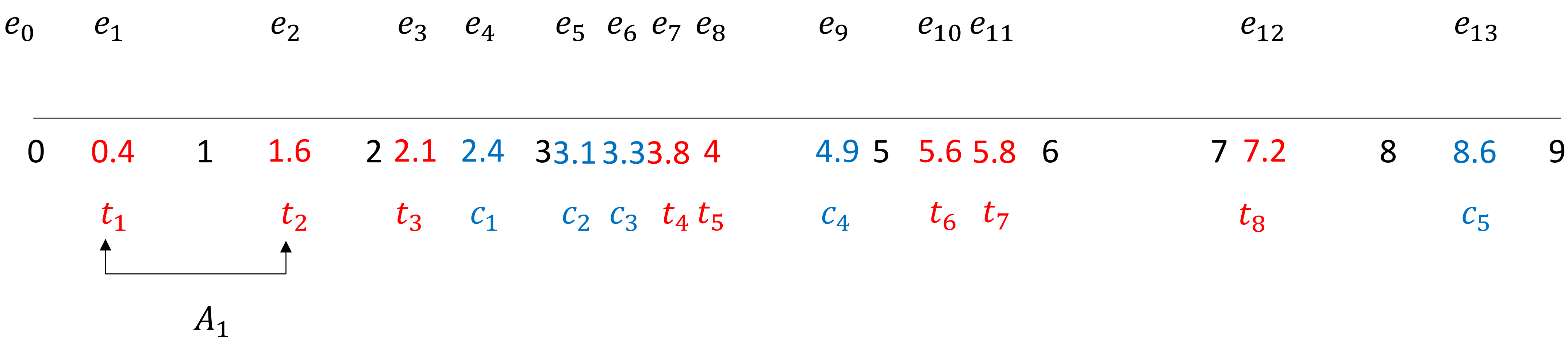
e_0 e_1 e_2 e_3 e_4 e_5 e_6 e_7 e_8 e_9 e_{10} e_{11} e_{12} e_{13}

0 0.4 1 1.6 2 2.1 2.4 3 3.1 3.3 3.8 4 4.9 5 5.6 5.8 6 7 7.2 8 8.6 9

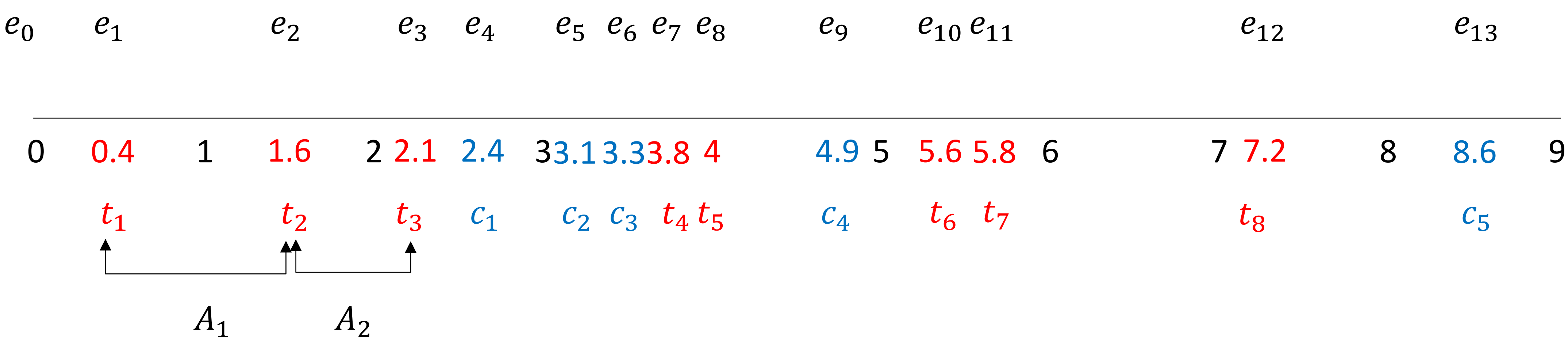
t_1 t_2 t_3 c_1 c_2 c_3 t_4 t_5 c_4 t_6 t_7 t_8 c_5



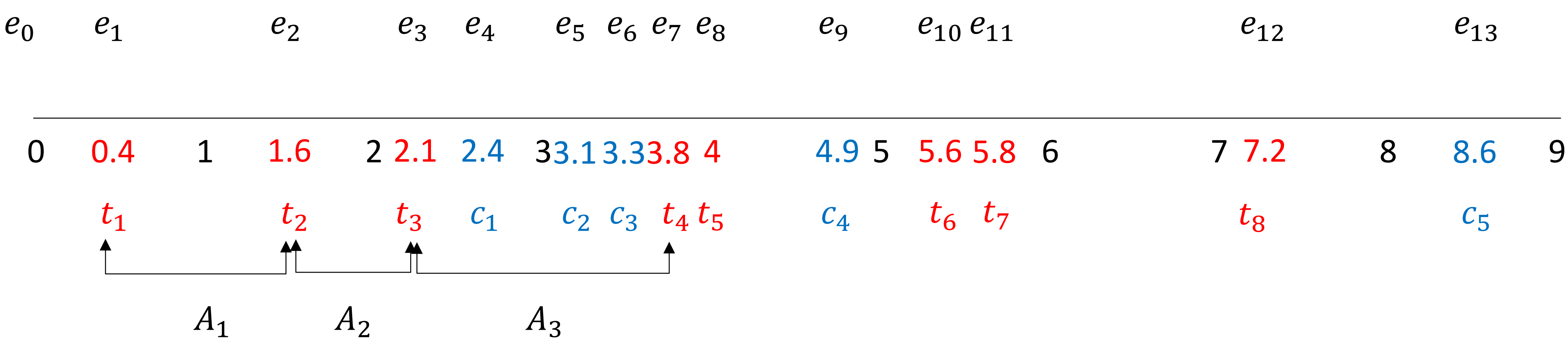
TIME BETWEEN ARRIVALS



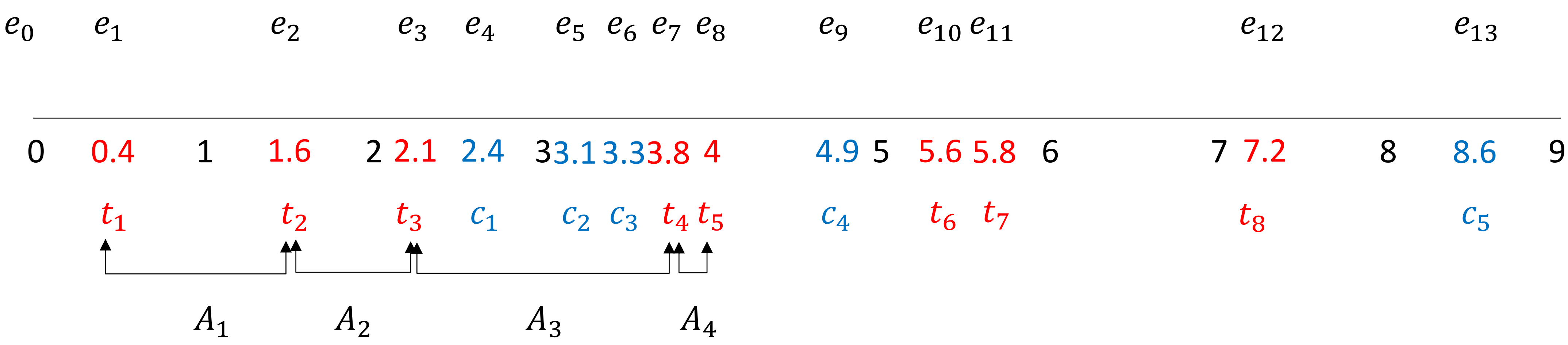
TIME BETWEEN ARRIVALS



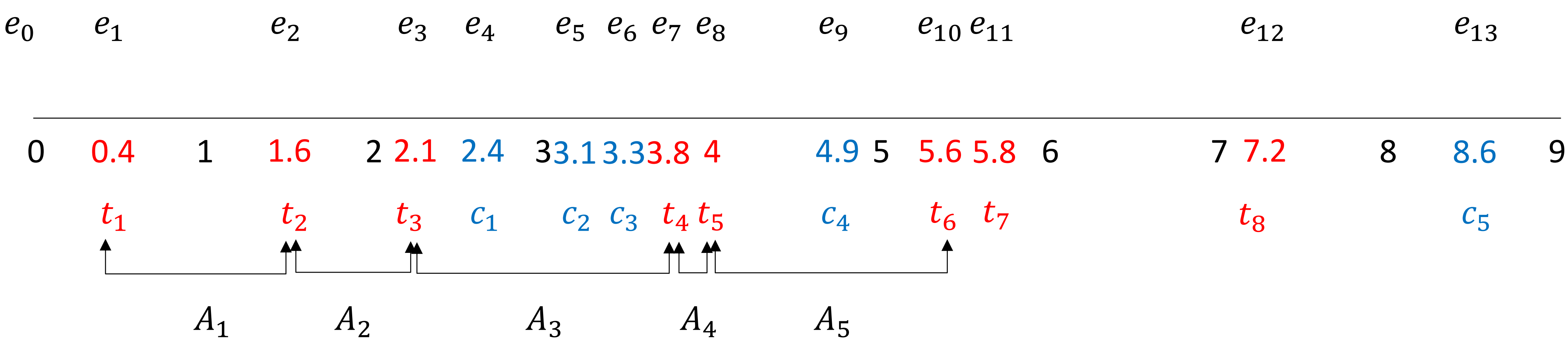
TIME BETWEEN ARRIVALS



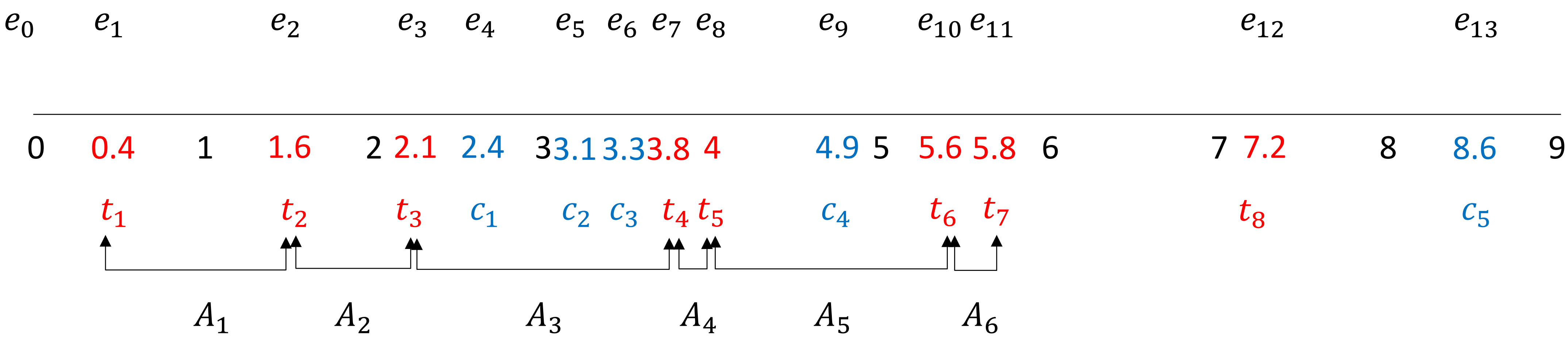
TIME BETWEEN ARRIVALS



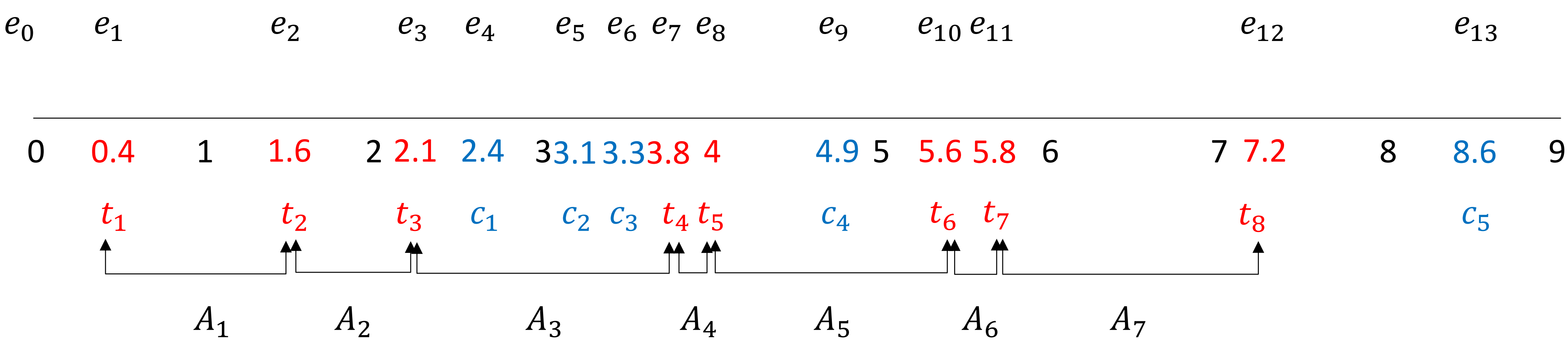
TIME BETWEEN ARRIVALS



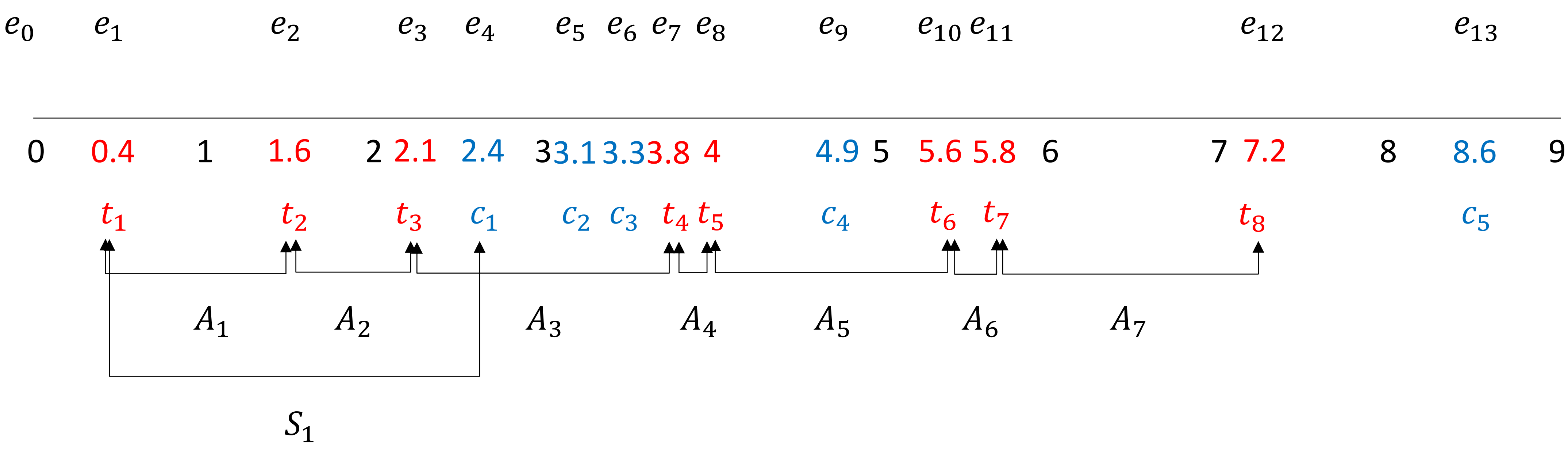
TIME BETWEEN ARRIVALS



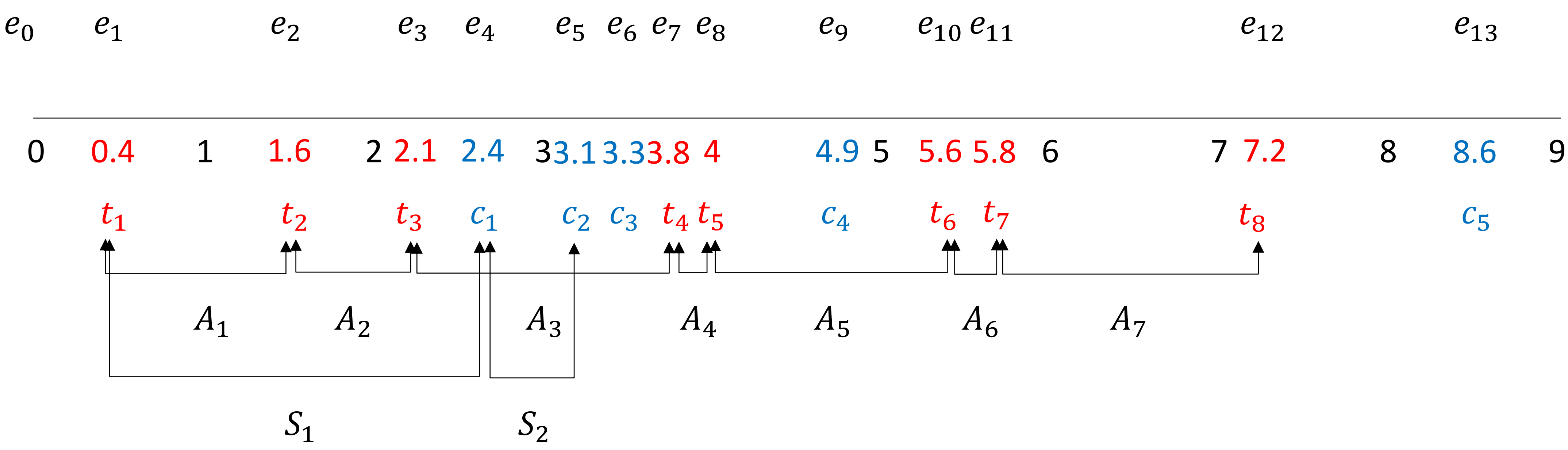
TIME BETWEEN ARRIVALS



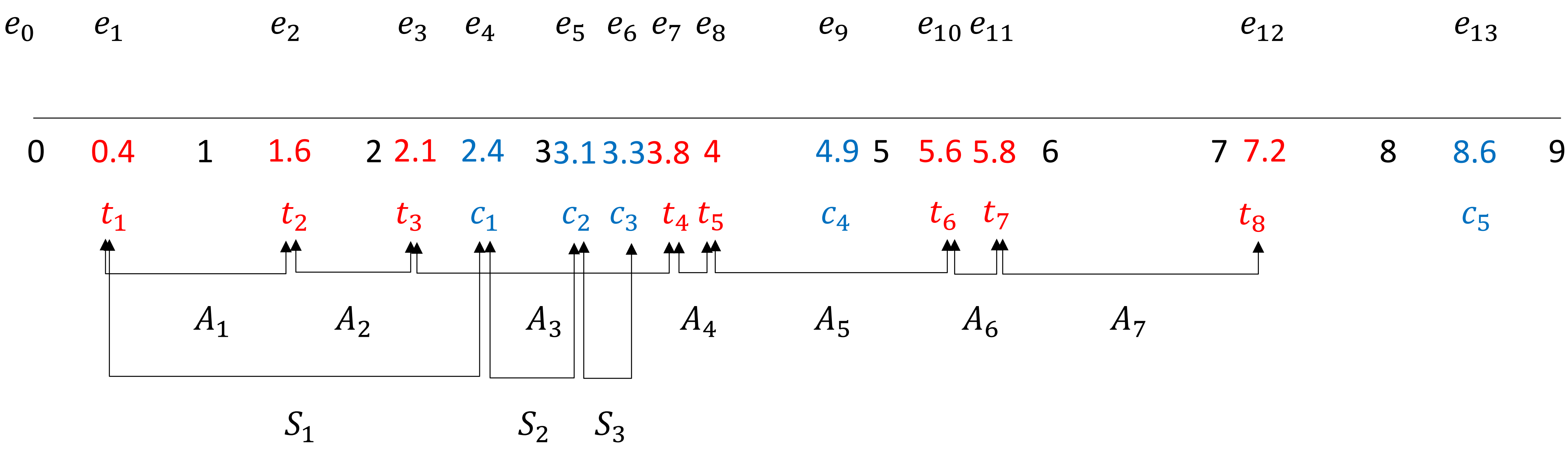
SERVICE TIME...



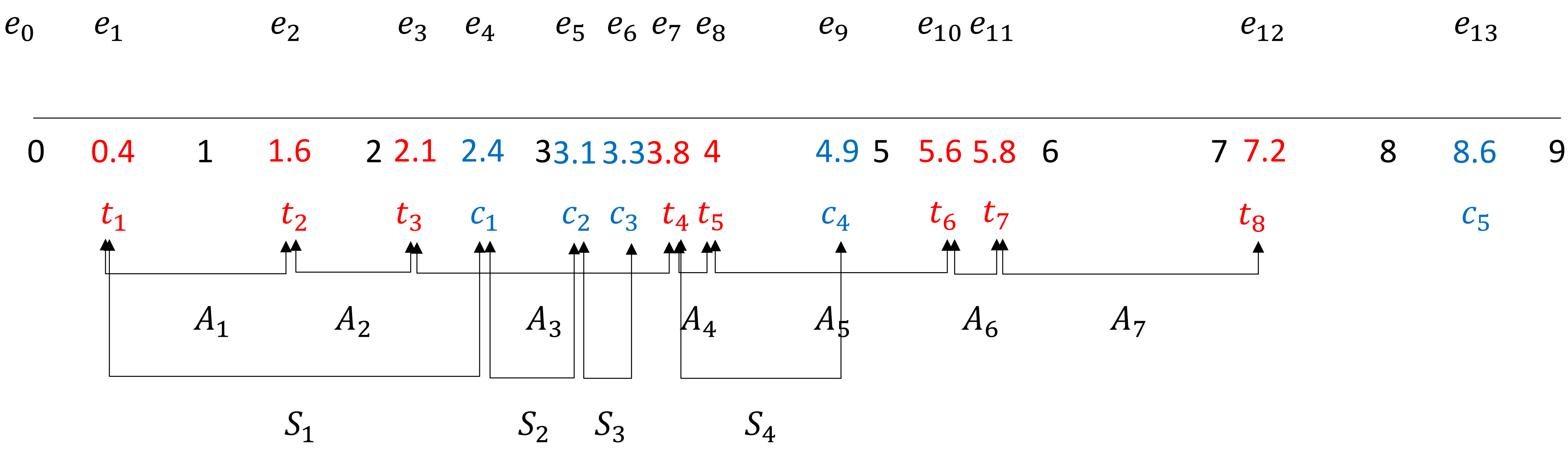
SERVICE TIME…



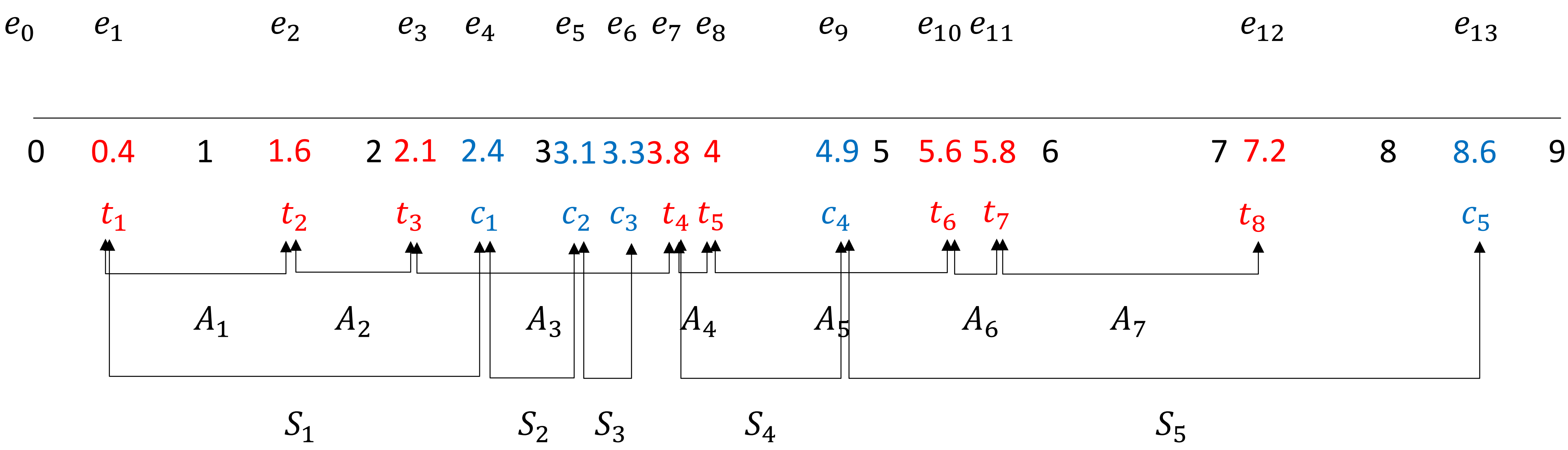
SERVICE TIME…



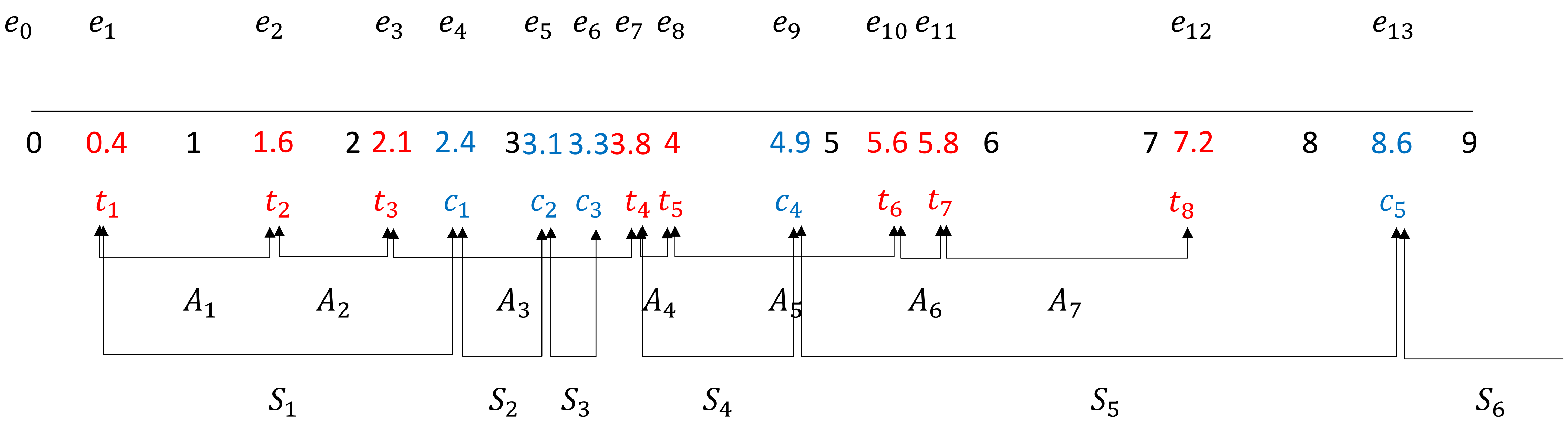
SERVICE TIME…



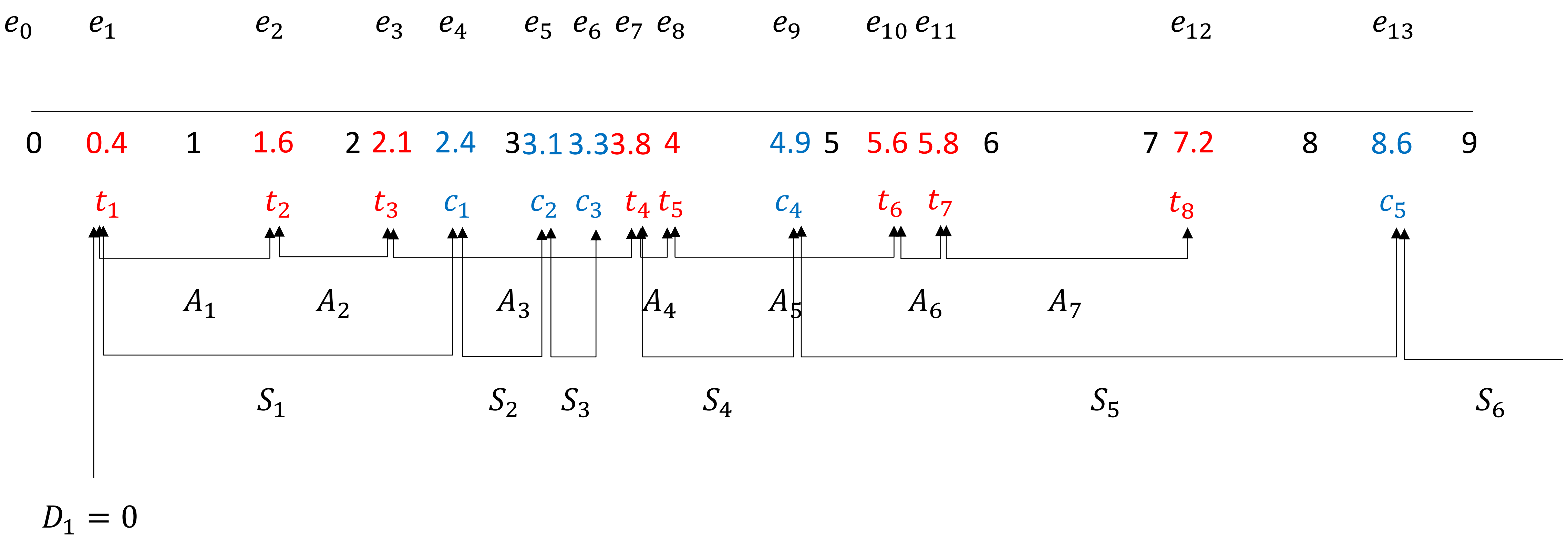
SERVICE TIME…



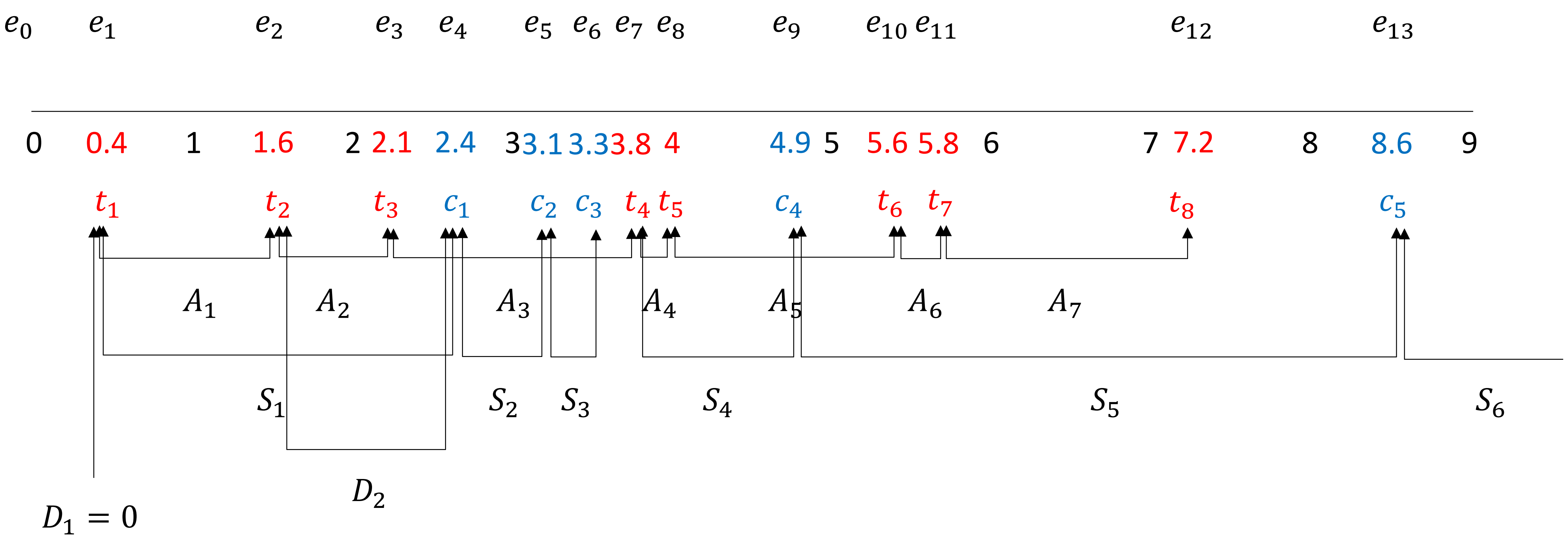
SERVICE TIME…



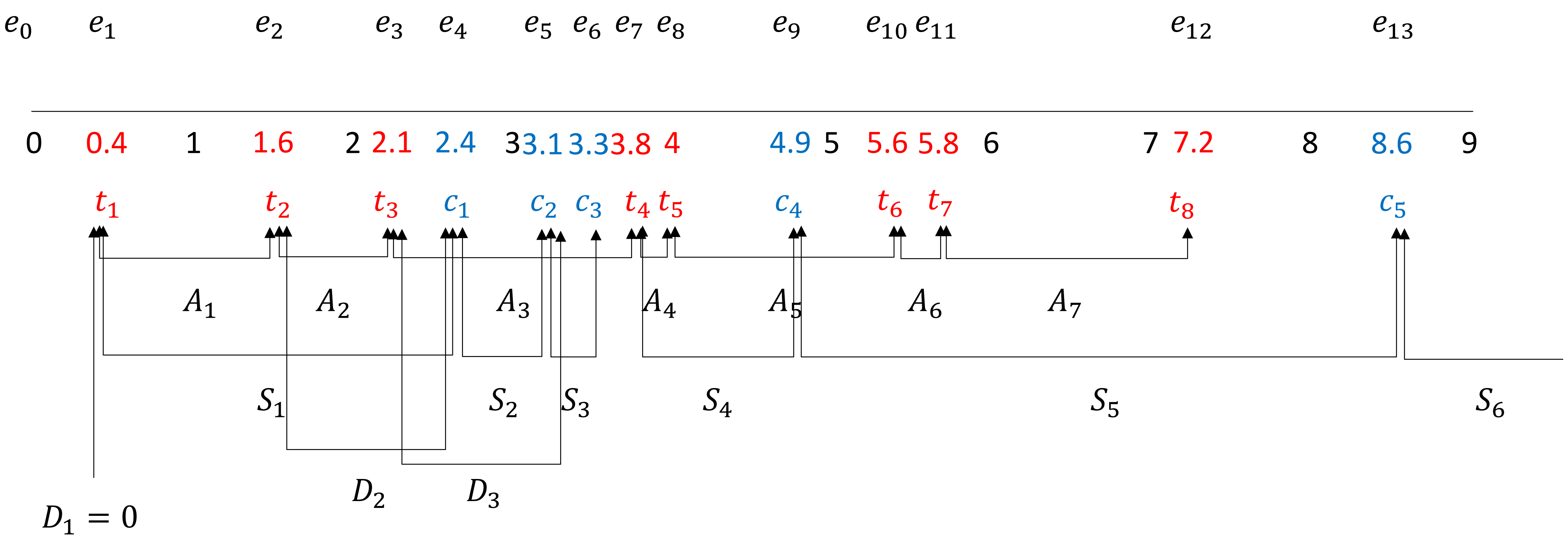
DELAY/WAITING TIME...



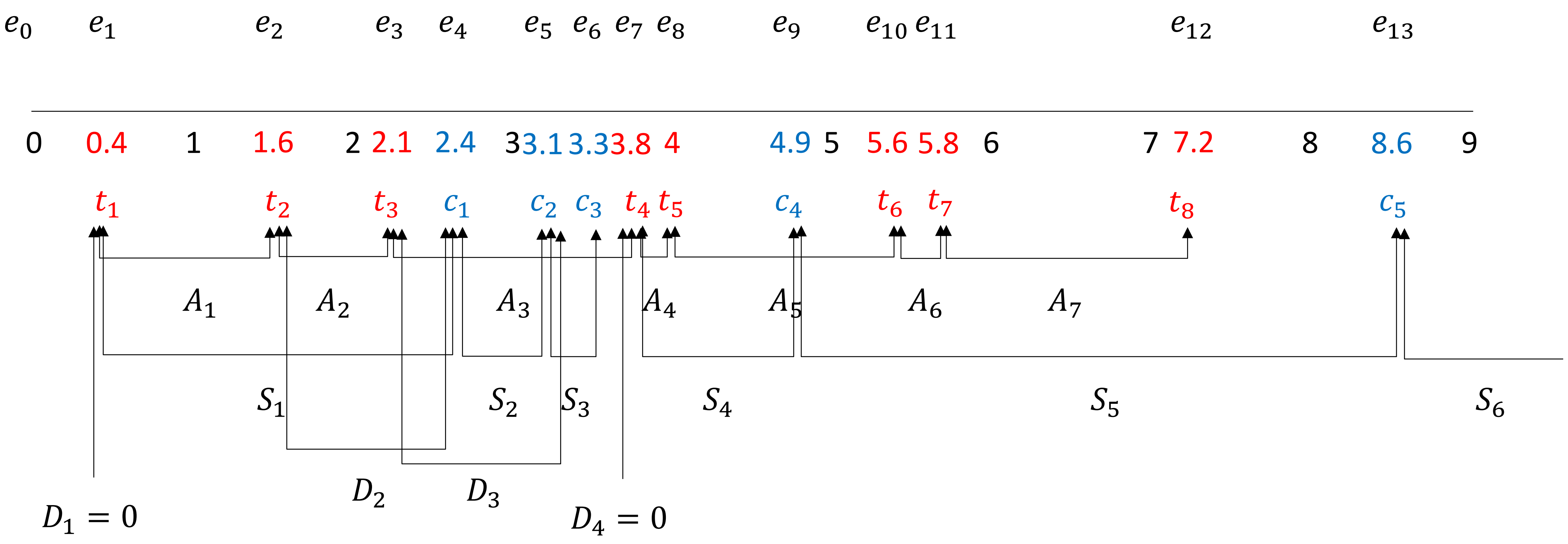
DELAY/WAITING TIME...



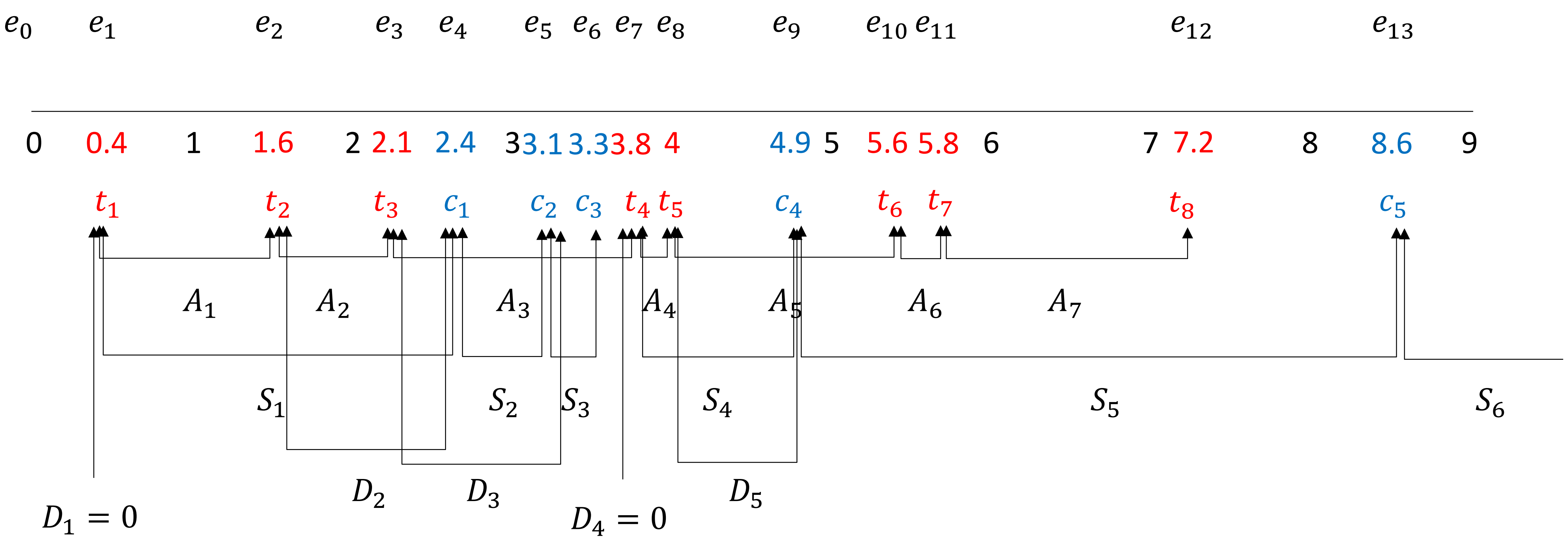
DELAY/WAITING TIME...



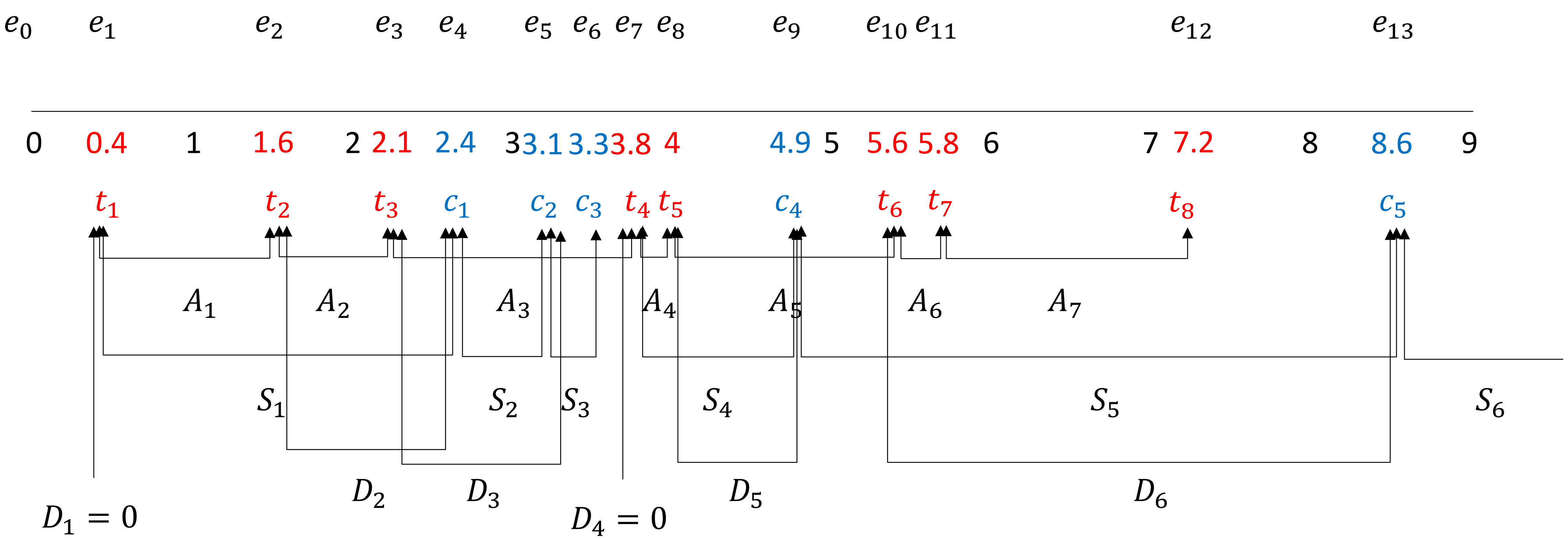
DELAY/WAITING TIME...



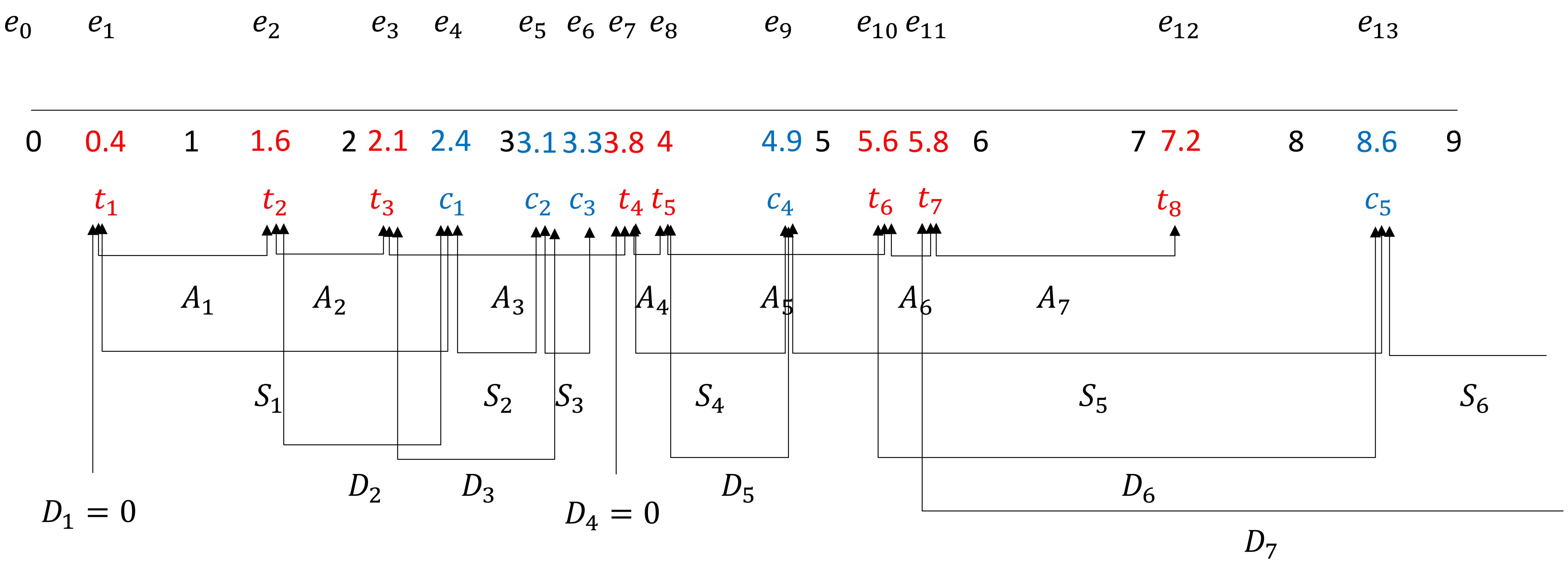
DELAY/WAITING TIME...



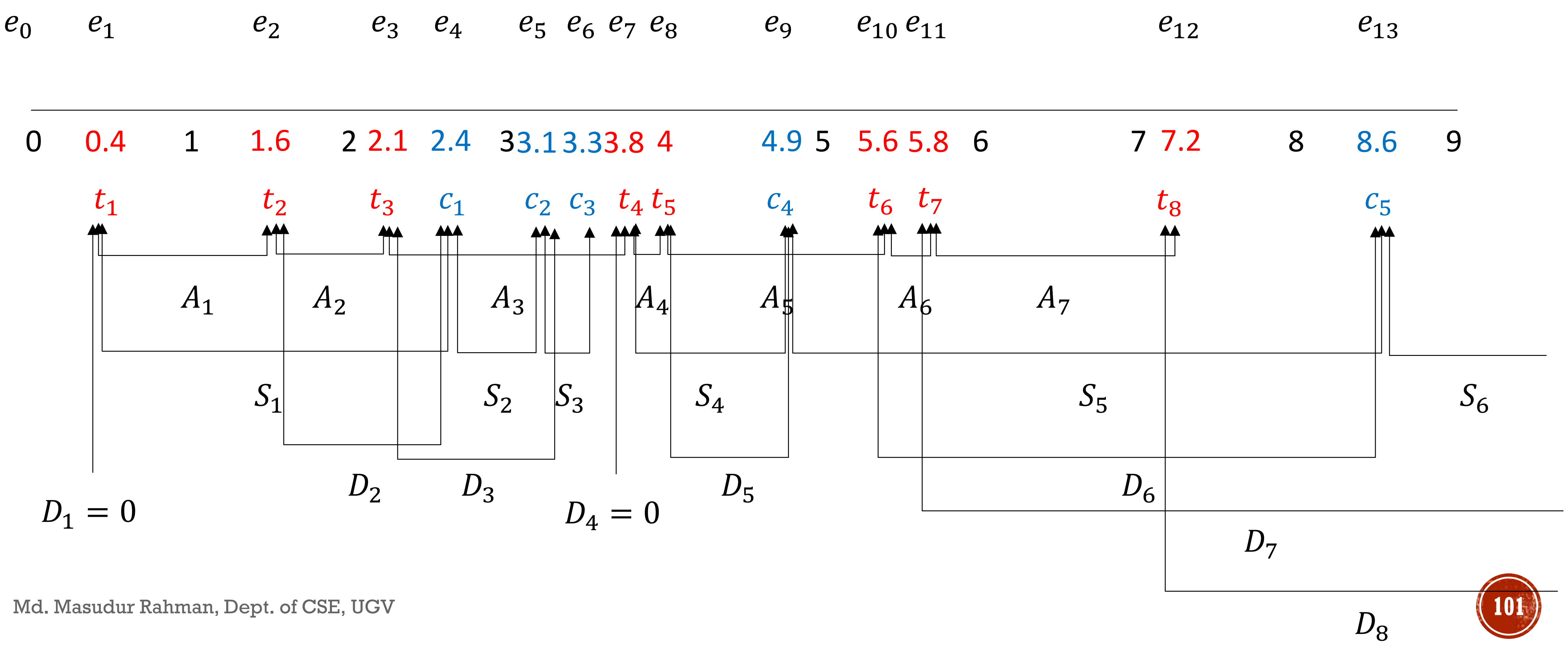
DELAY/WAITING TIME…



DELAY/WAITING TIME...



DELAY/WAITING TIME...



WEEK 9

SLIDES 102-123

103

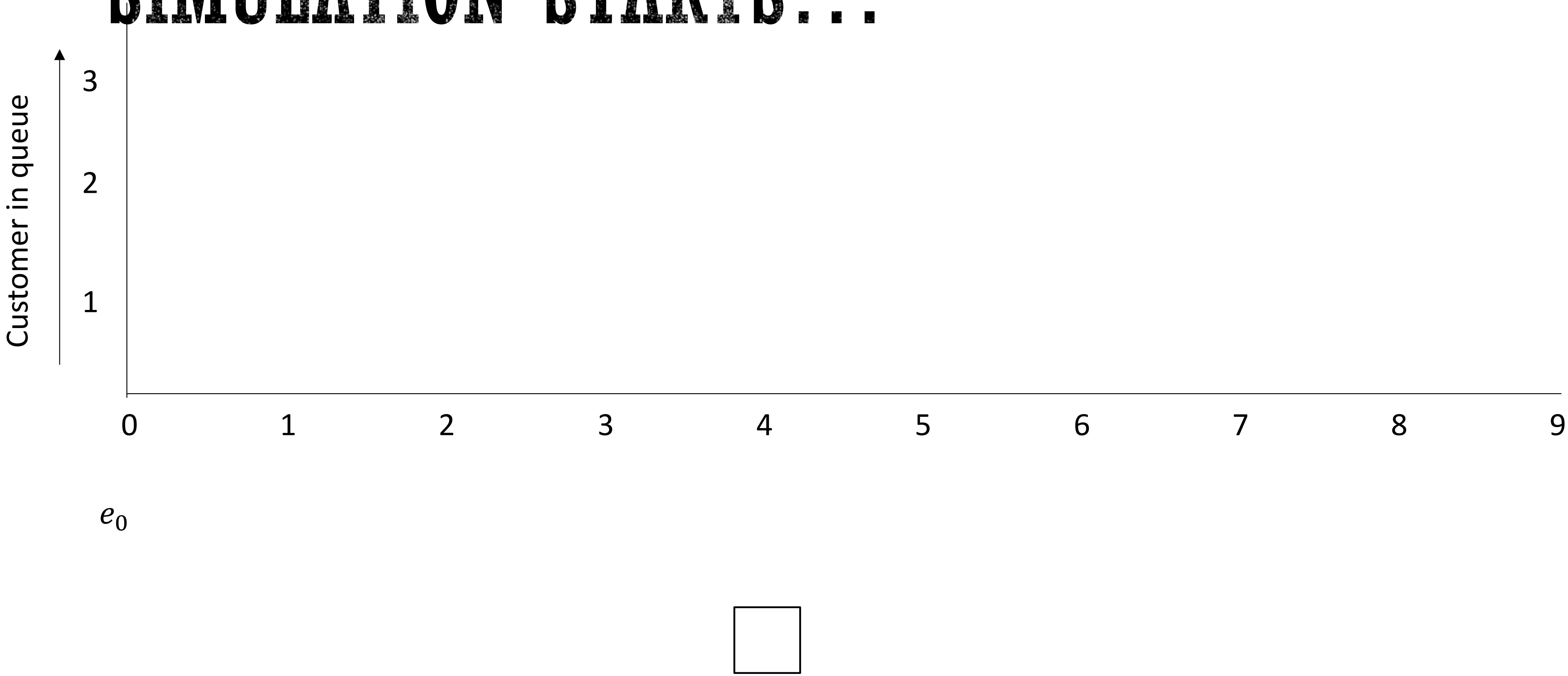
PERFORMANCE EVALUATION

Md. Masudur Rahman, Dept. of CSE, UGV

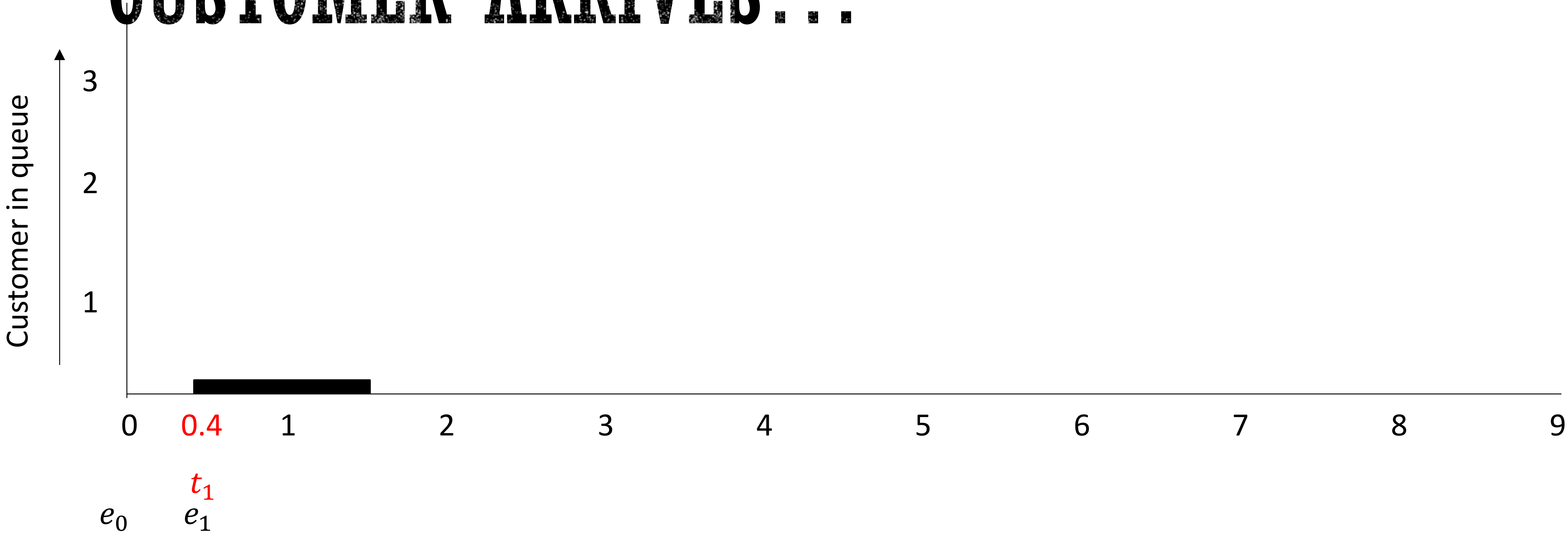
CALCULATION OF AVERAGE QUEUE LENGTH

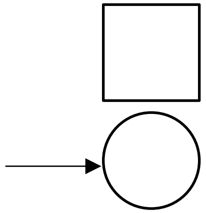
- Average queue length, $\hat{q}(n) = \frac{\sum_{i=0}^{\infty} iT_i}{T(n)}$
- T_i is the time length that i number of customers spend in queue
- $T(n)$ is the time when the simulation ends
- Let $T(n) = 8.6$

SIMULATION STARTS…

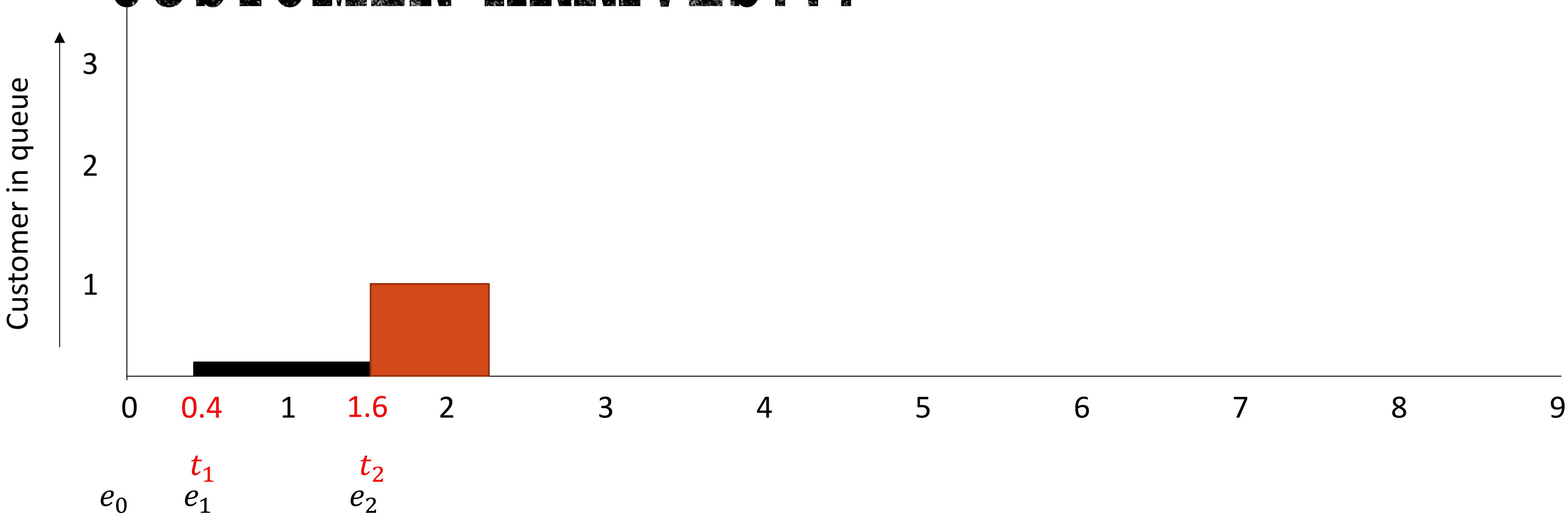


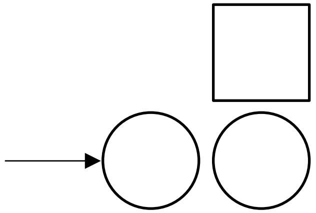
CUSTOMER ARRIVES...



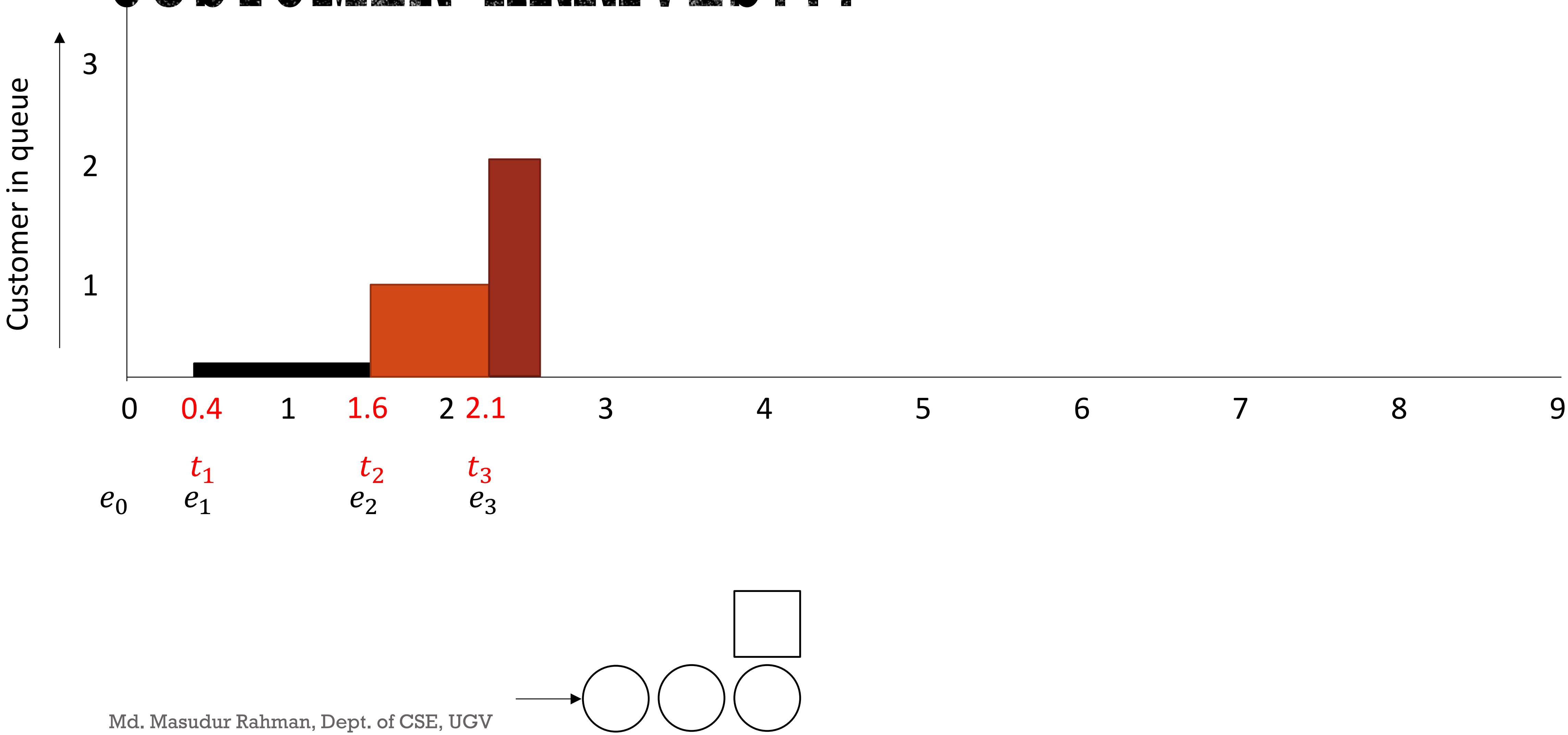


CUSTOMER ARRIVES...

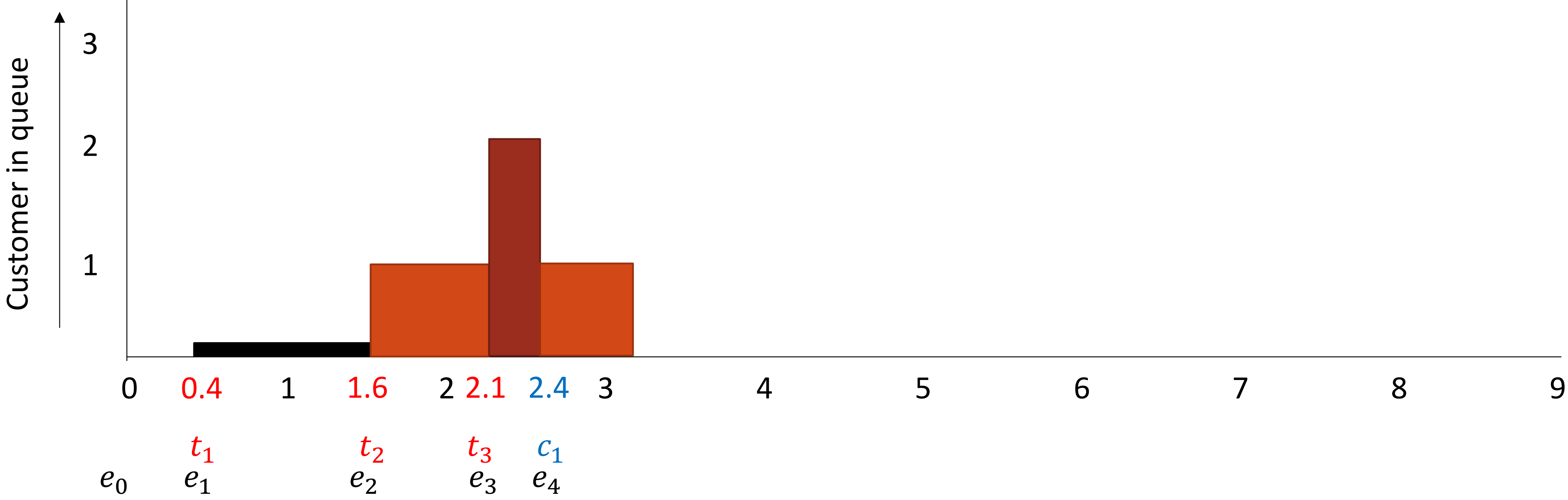


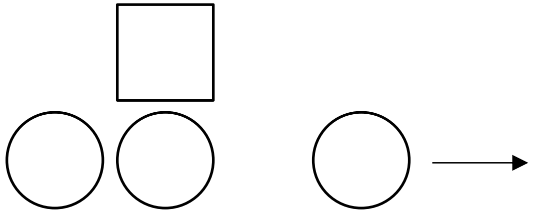


CUSTOMER ARRIVES...

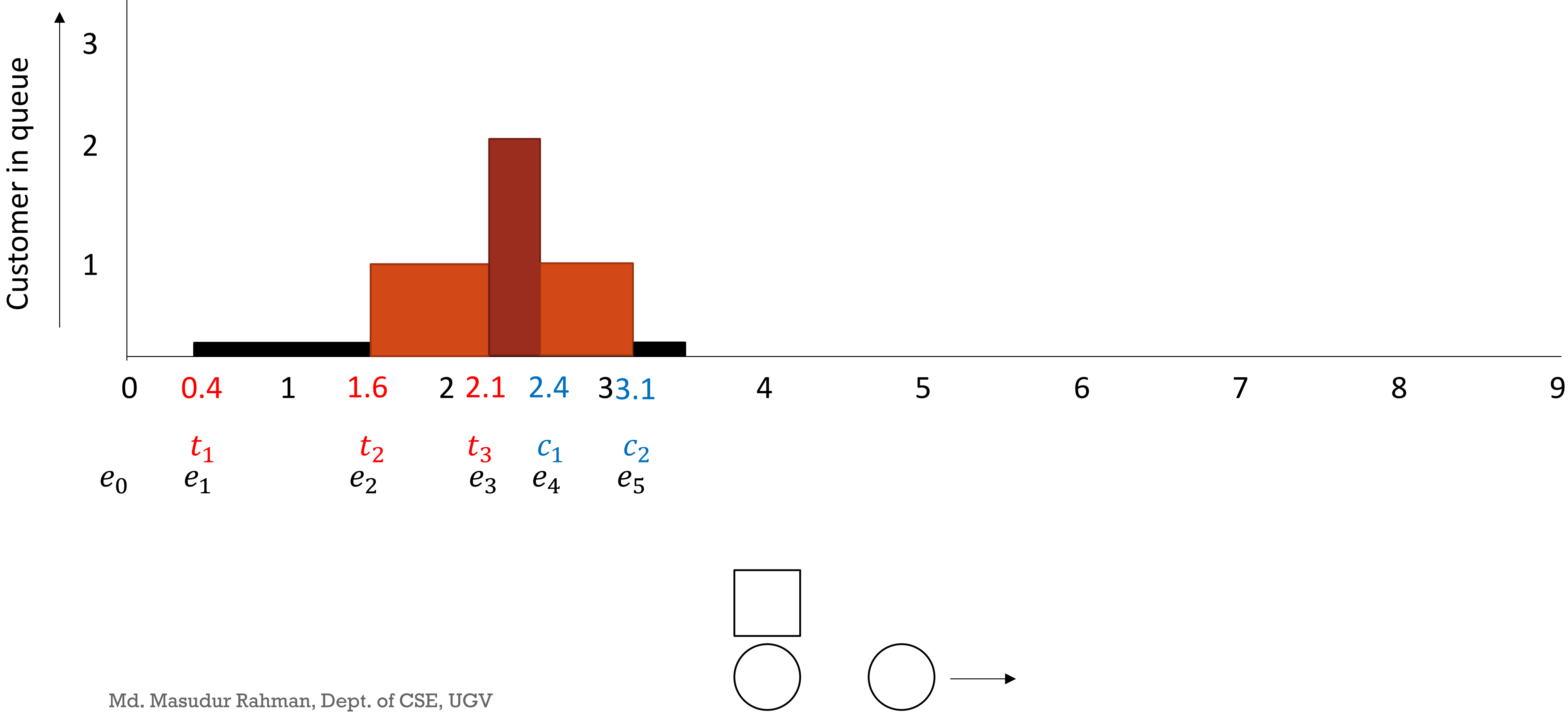


CUSTOMER LEAVES...

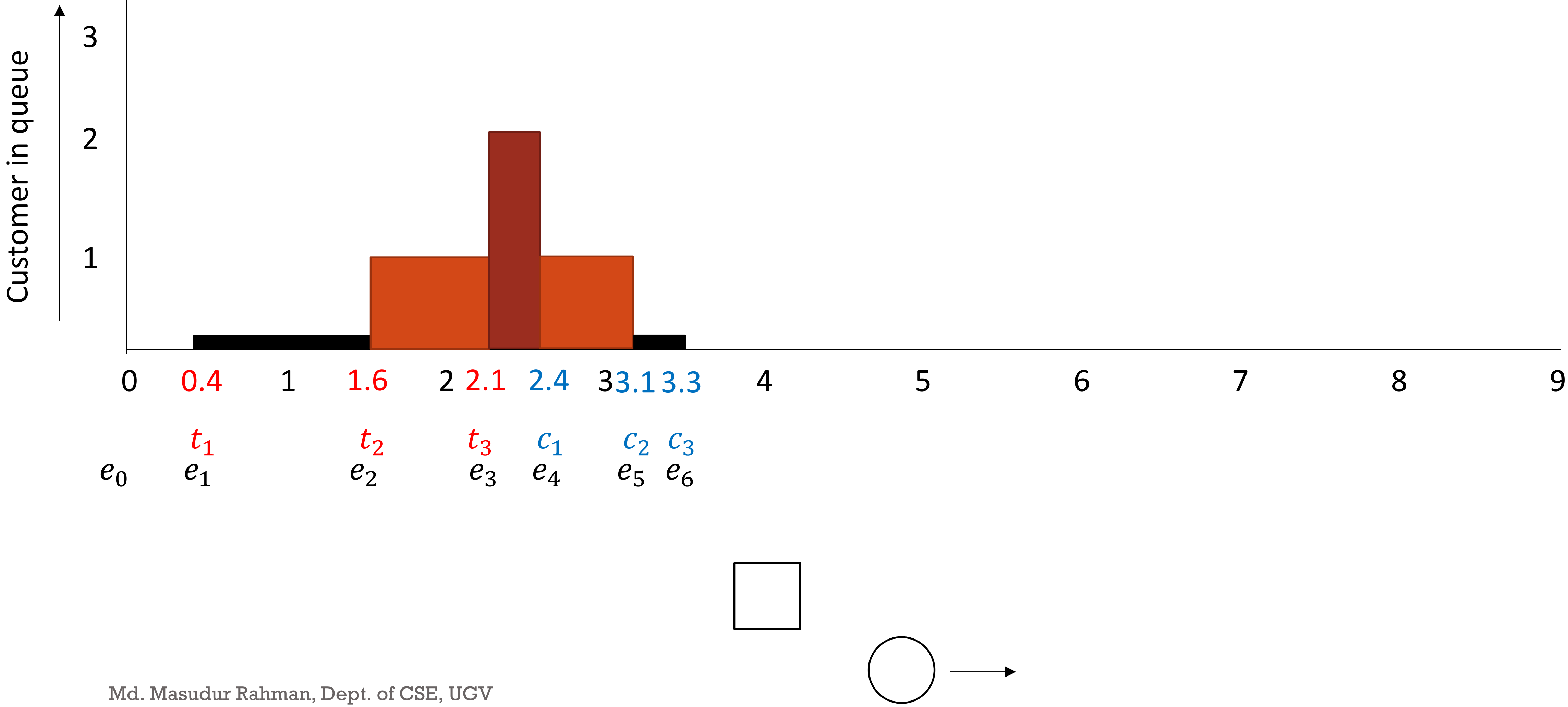




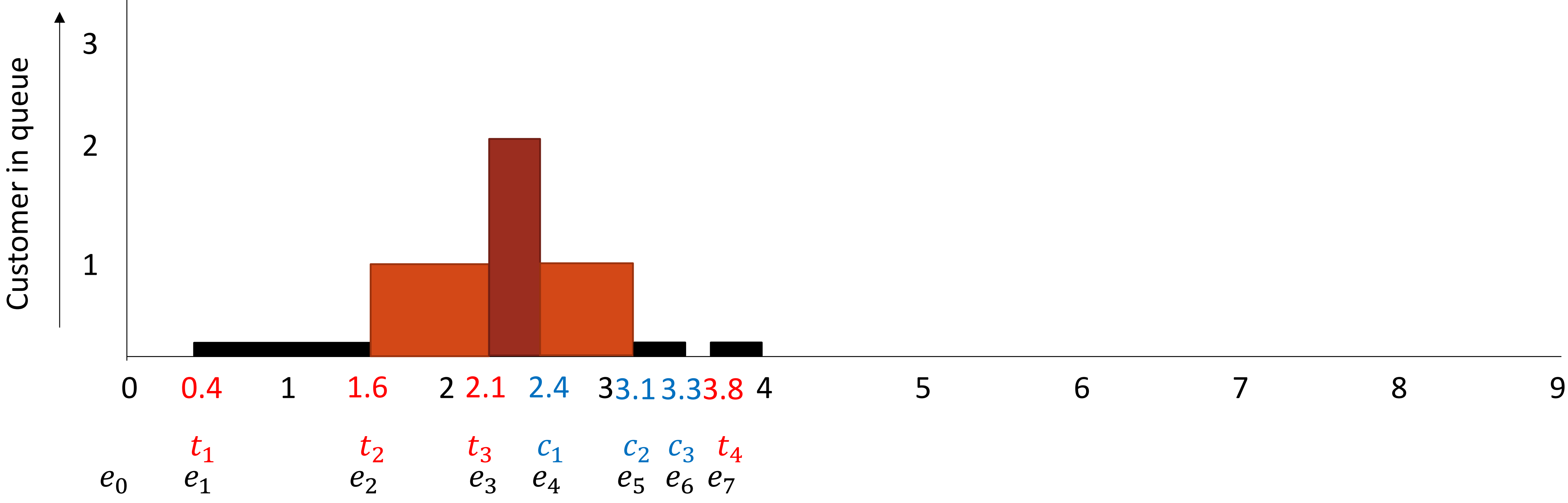
CUSTOMER LEAVES…

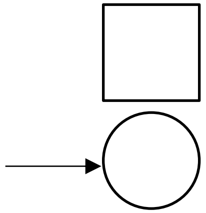


CUSTOMER LEAVES...

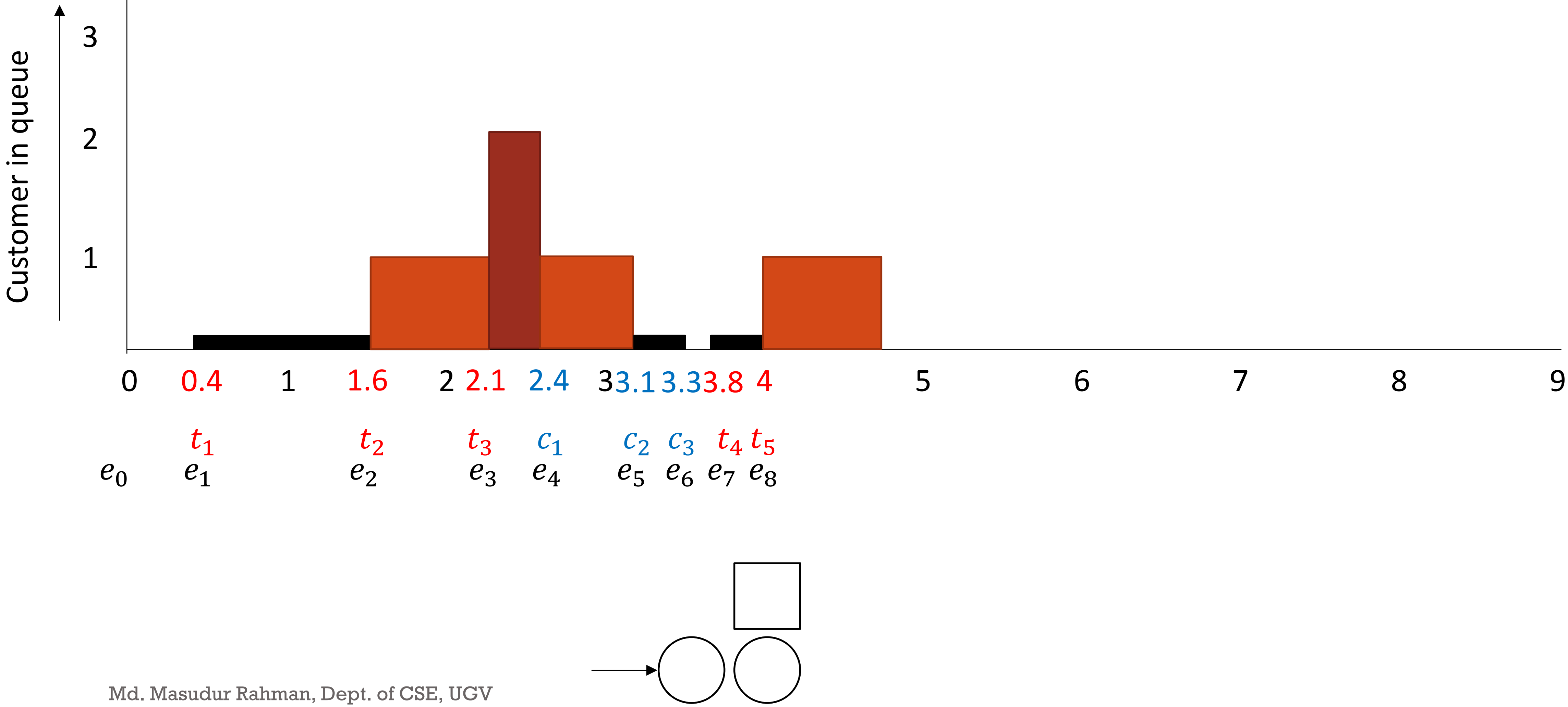


CUSTOMER ARRIVES...

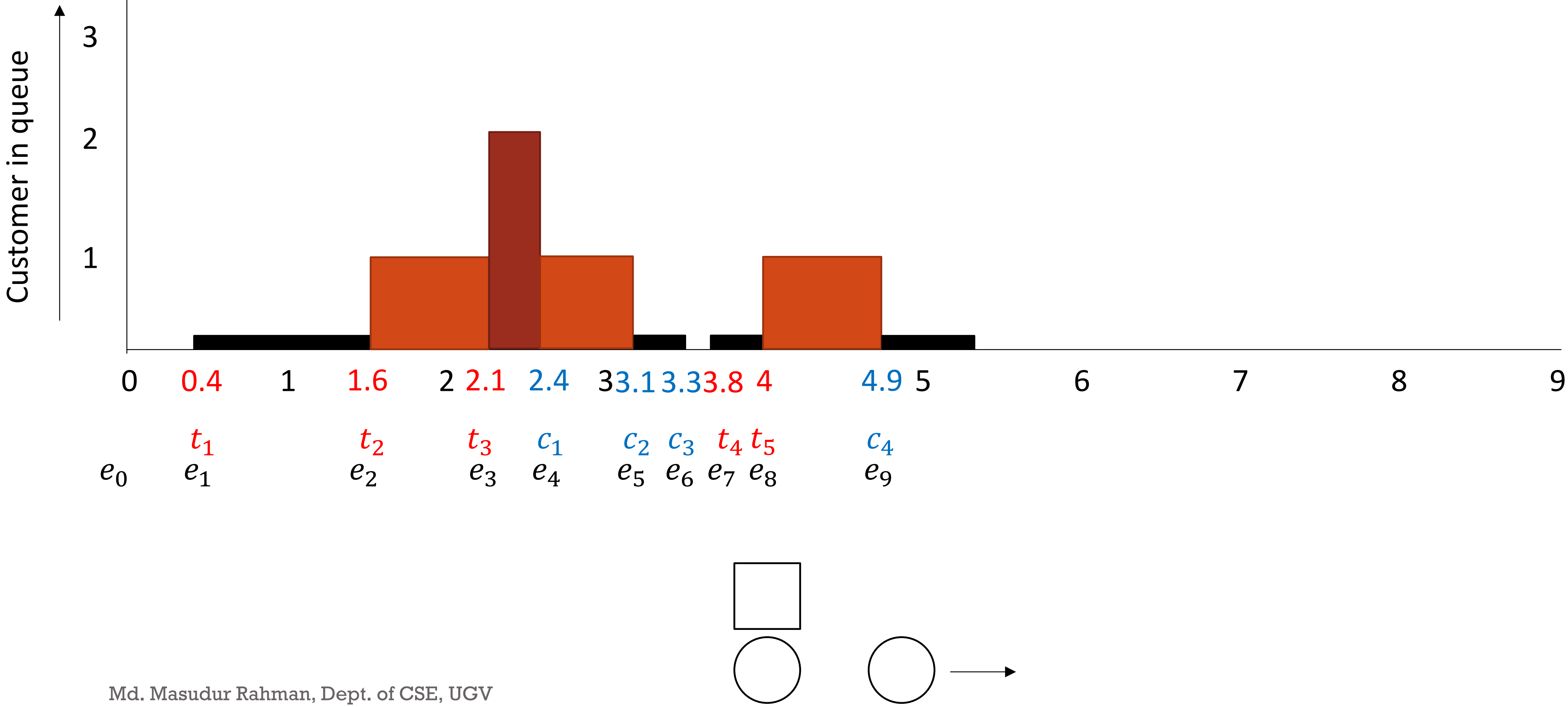




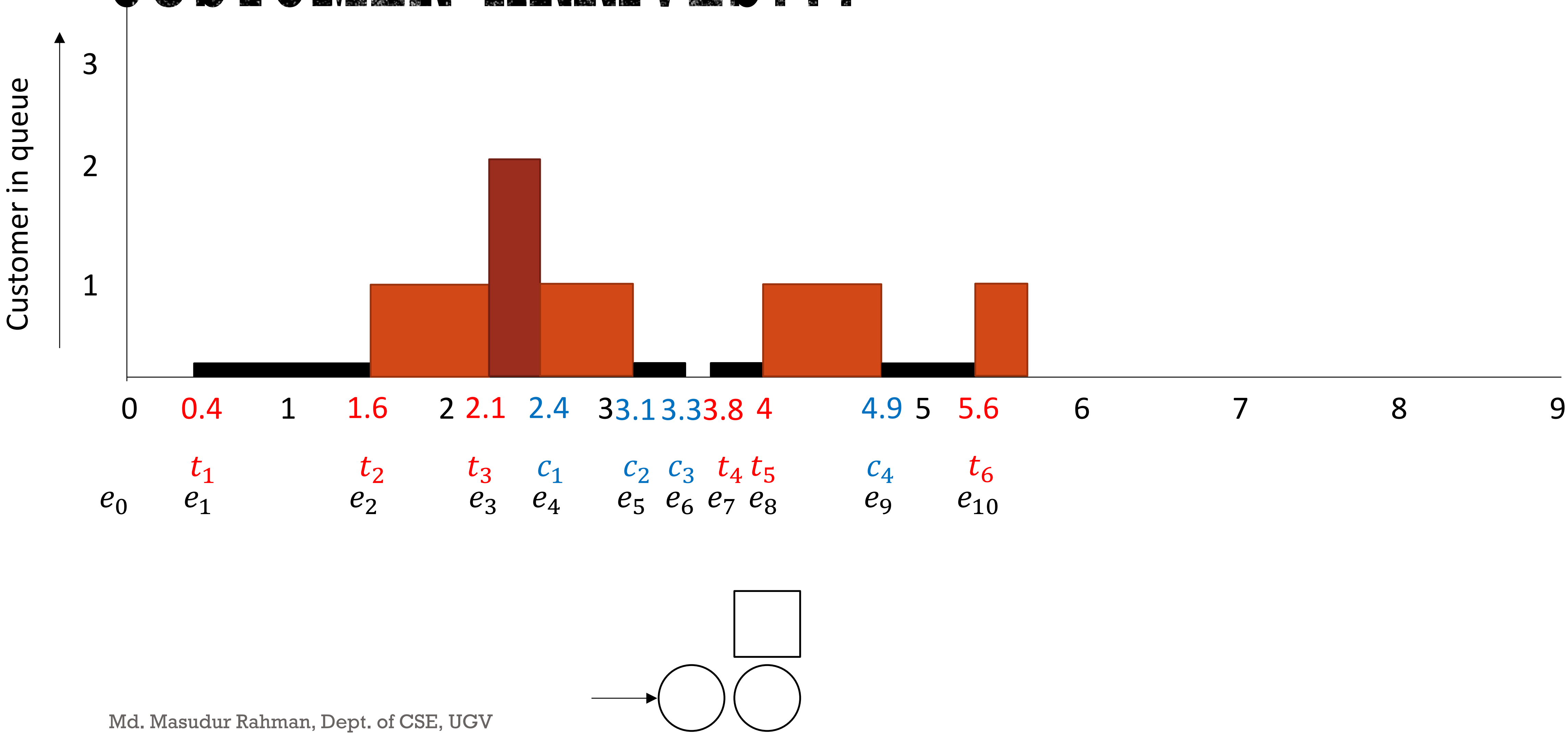
CUSTOMER ARRIVES…



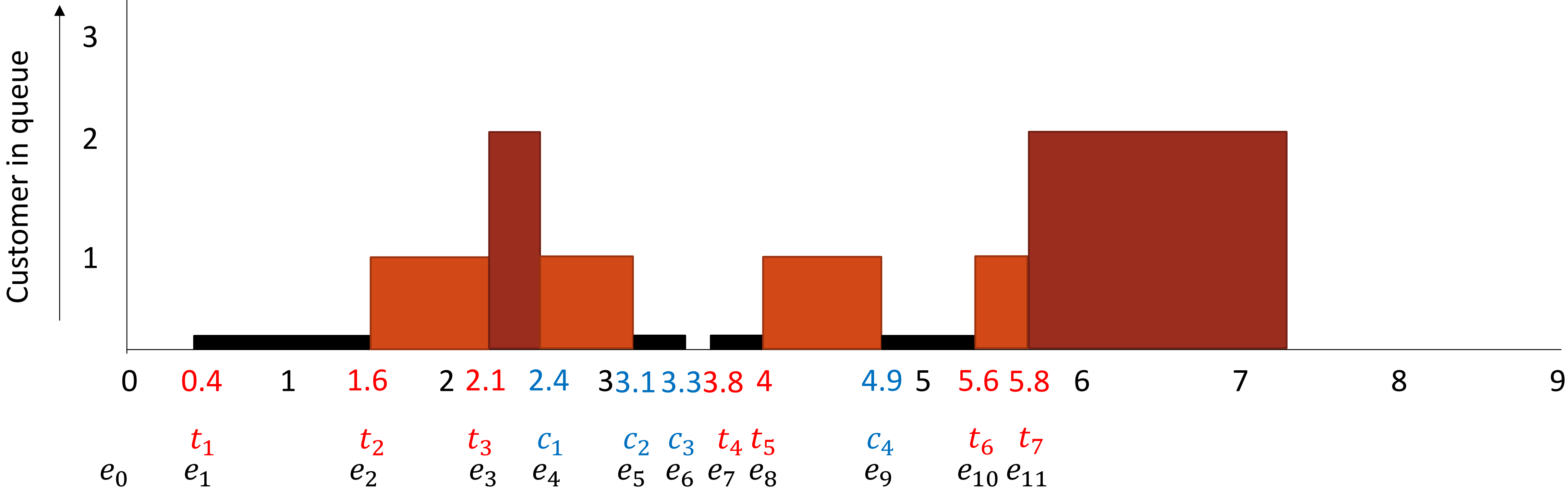
CUSTOMER LEAVES...

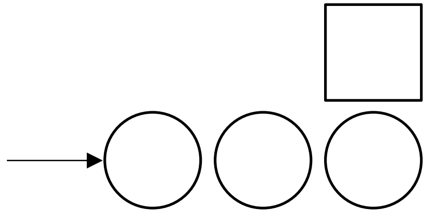


CUSTOMER ARRIVES…

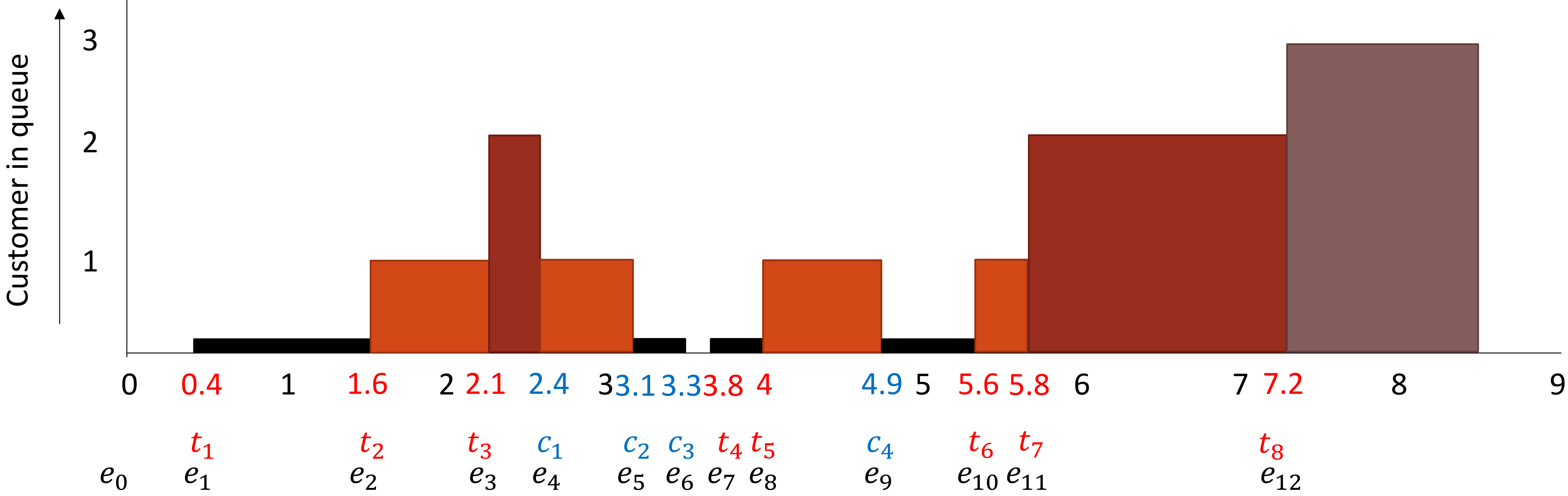


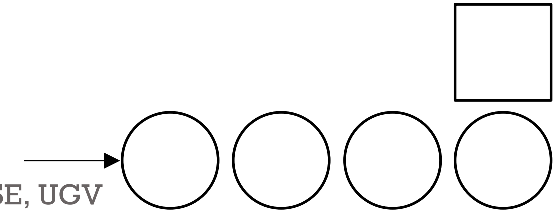
CUSTOMER ARRIVES…



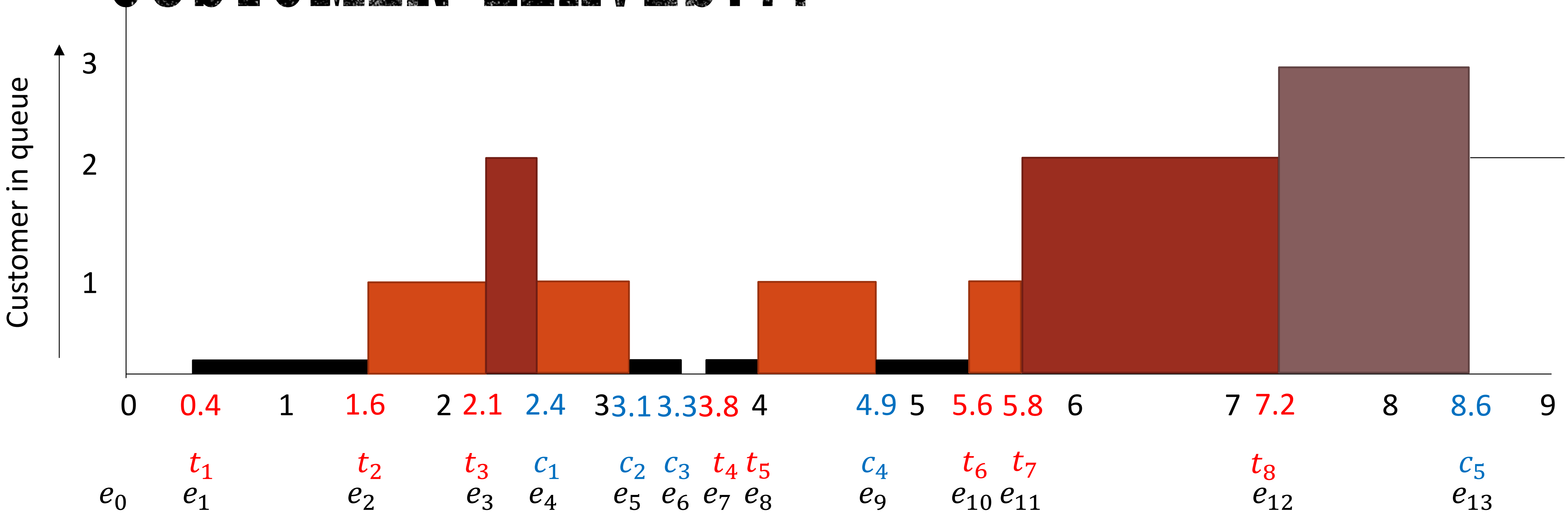


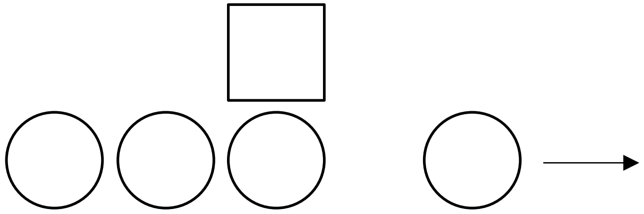
CUSTOMER ARRIVES...



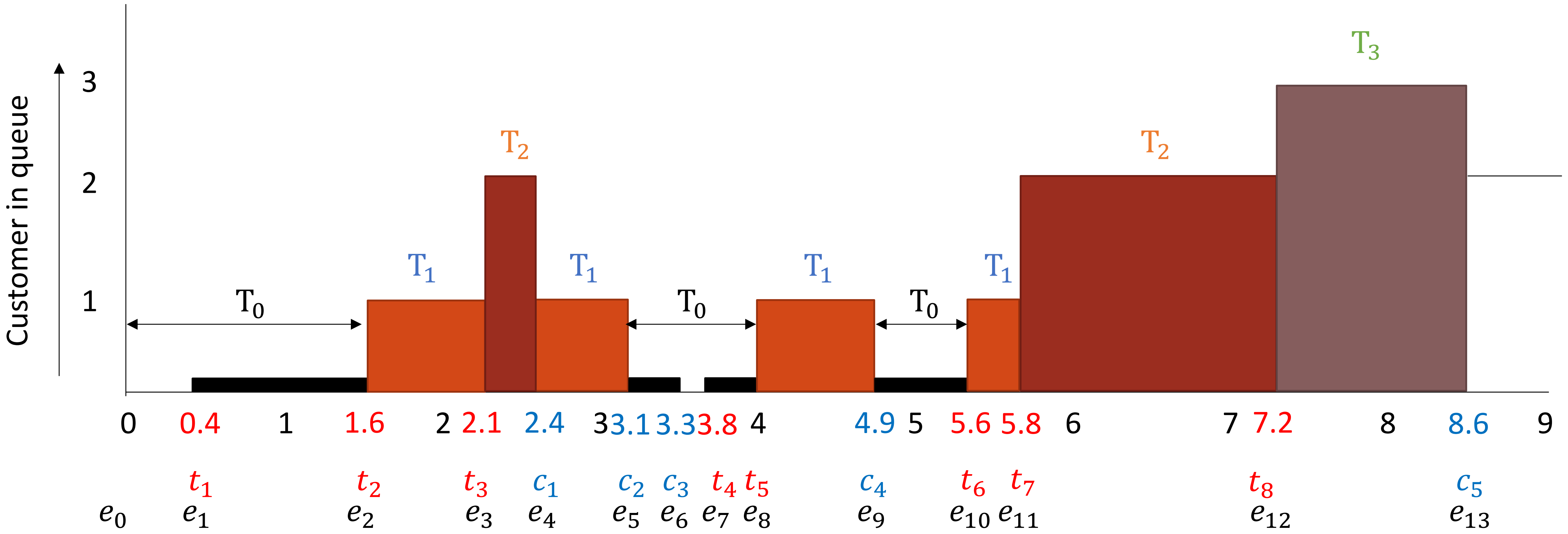


CUSTOMER LEAVES…





CALCULATION OF T_i



- $T_0 = (1.6 - 0.0) + (4.0 - 3.1) + (5.6 - 4.9) = 3.2$
- $T_1 = (2.1 - 1.6) + (3.1 - 2.4) + (4.9 - 4.0) + (5.8 - 5.6) = 2.3$
- $T_2 = (2.4 - 2.1) + (7.2 - 5.8) = 1.7$
- $T_3 = (8.6 - 7.2) = 1.4$

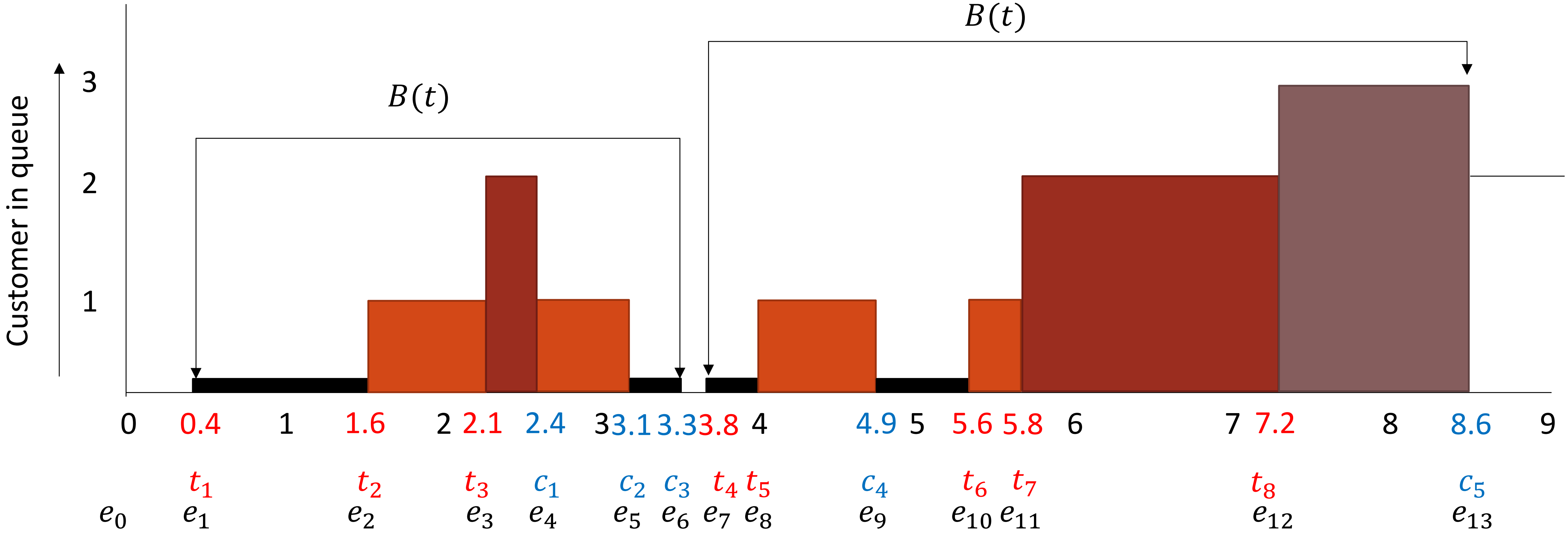
CALCULATION OF $\sum_{i=0}^{\infty} iT_i$ AND $\hat{q}(n)$

- $\sum_{i=0}^{\infty} iT_i = 0 \times T_0 + 1 \times T_1 + 2 \times T_2 + 3 \times T_3 + \cdots$
- $\sum_{i=0}^{\infty} iT_i = 0 \times 3.2 + 1 \times 2.3 + 2 \times 1.7 + 3 \times 1.4 = 9.9$
- $\hat{q}(n) = \frac{\sum_{i=0}^{\infty} iT_i}{T(n)} = \frac{9.9}{8.6} = 1.15$

CALCULATION OF UTILIZATION TIME

- Utilization time, $\hat{u}(n) = \frac{B(t)}{T(n)}$
- $B(t)$ = Time length for which the server was busy
- $T(n)$ = Simulation end time = 8.6

CALCULATION OF $B(t)$



- $B(t) = (3.3 - 0.4) + (8.6 - 3.8) = 7.7$

CALCULATION OF $\hat{u}(n)$

- $\hat{u}(n) = \frac{B(t)}{T(n)} = \frac{7.7}{8.6} = 0.90$

WEEK 10

SLIDES 124-130

SIMULATION OF INVENTORY SYSTEM

Md. Masudur Rahman, Dept. of CSE, UGV

SIMULATION OF INVENTORY SYSTEM

- A company that sells a single product would like to decide how many items it should have in inventory for each of the next *n* months (*n* is a fixed input parameter). *D* denotes the *demand size* for the products. (w.p: with probability).

$$D = \begin{cases} 1 & w.p\ \frac{1}{6} \\ 2 & w.p\ \frac{1}{3} \\ 3 & w.p\ \frac{1}{3} \\ 4 & w.p\ \frac{1}{6} \end{cases}$$

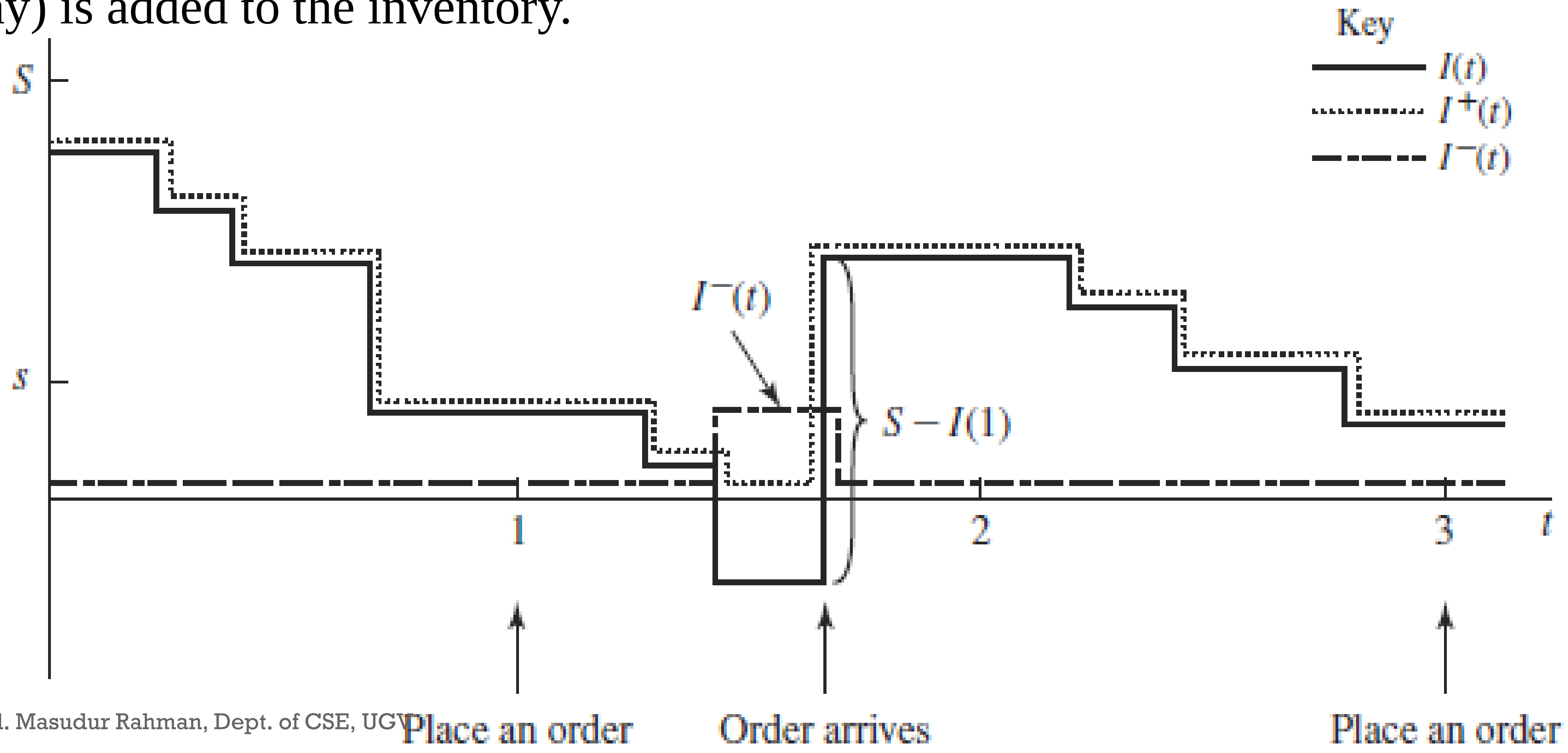
SIMULATION OF INVENTORY SYSTEM

- If the company orders Z items, it incurs a cost of $K + iZ$, where $K = \$32$ is the *setup cost* and $i = \$3$ is the *incremental cost* per item ordered. (If $Z = 0$, no cost is incurred.) When an order is placed, the time required for it to arrive (called the *delivery lag* or *lead time*) is a random variable that is distributed uniformly between 0.5 and 1 month.
- The company uses a stationary (s, S) policy to decide how much to order, i.e.,

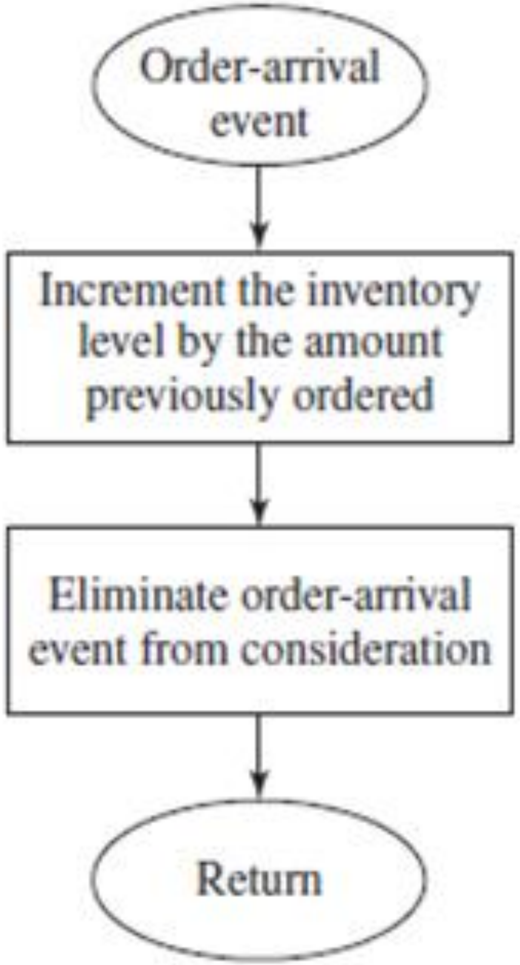
$$Z = \begin{cases} S - I & if\ I < s \\ 0 & if\ I \geq s \end{cases}$$

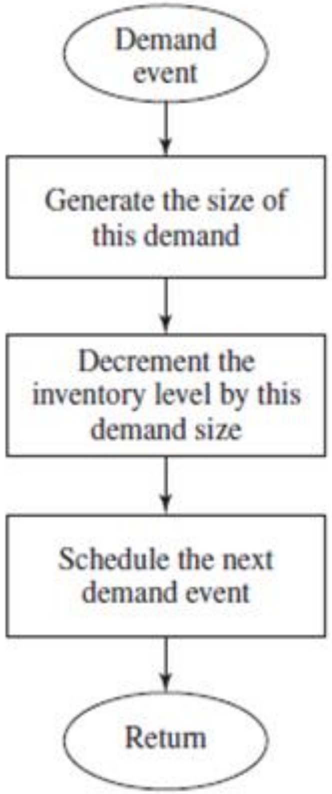
where I is the inventory level at the beginning of the month. I^+ denotes holding and I^- denotes shortage.

- When a demand occurs, it is satisfied immediately if the inventory level is at least as large as the demand. If the demand exceeds the inventory level, the excess of demand over supply is backlogged and satisfied by future deliveries. (In this case, the new inventory level is equal to the old inventory level minus the demand size, resulting in a negative inventory level.) When an order arrives, it is first used to eliminate as much of the backlog (if any) as possible; the remainder of the order (if any) is added to the inventory.

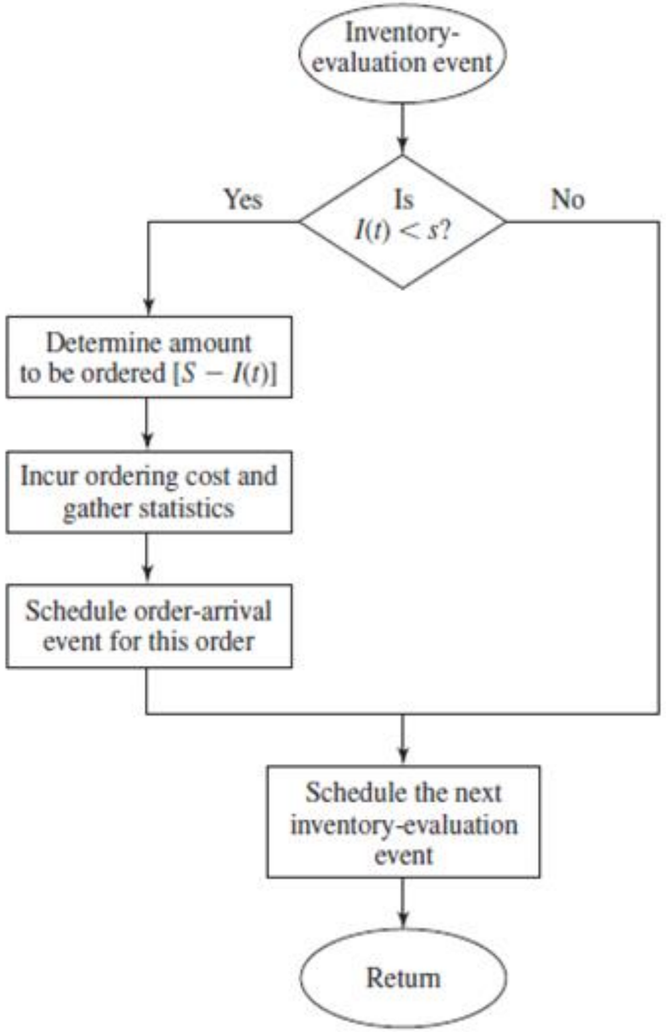


FLOWCHART FOR ORDER ARRIVAL AND DEMAND





FLOWCHART FOR INVENTORY EVALUATION



WEEK 11

SLIDES 131-139

132

RANDOM VARIABLES

EXPERIMENT, SAMPLE SPACE, SAMPLE POINTS

- An experiment is a process whose outcome is not known with certainty. The set of all possible outcomes of an experiment is called the sample space and is denoted by *S*. The outcomes themselves are called the sample points in the sample space.

- **Example 1:** If the experiment consists of flipping a coin, then

$$S = \{H, T\}$$

where the symbol {} means the "set consisting of," and "H" and "T" mean that the outcome is a head and a tail, respectively.

- **Example 2:** If the experiment consists of tossing a die, then

$$S = \{1, 2, \ldots, 6\}$$

where the outcome i means that i appeared on the die, $i = 1, 2, \ldots, 6$.

RANDOM VARIABLE

- A random variable is a function (or rule) that assigns a real number (any number greater than $-\infty$ and less than ∞) to each point in the sample space S.

- **Example 3:** Consider the experiment of rolling a pair of dice. Then

$$S = \{(1,1),(1,2),\ldots,(6,6)\}$$

where (i,j) means that i and j appeared on the first and second die, respectively. If X is the random variable corresponding to the sum of the two dice, then X assigns the value 7 to the outcome (4, 3).

- **Example 4:** Consider the experiment of flipping two coins. If X is the random variable corresponding to the number of heads that occur, then X assigns the value 1 to either the outcome (H, T) or the outcome (T, H).

- **Denoting Random Variables:** In general, we denote random variables by capital letters such as X, Y, Z and the values that random variables take on by lowercase letters such as x, y, z.

DISTRIBUTION FUNCTION

- The distribution function (sometimes called the cumulative distribution function) F(x) of the random variable X is defined for each real number x as follows:

$$F(x) = P(X \leq x) \quad \text{for} -\infty < x < \infty$$

- where $P(X \leq x)$ means the probability associated with the event $\{X \leq x\}$. Thus, $F(x)$ is the probability that, when the experiment is done, the random variable X will have taken on a value no larger than the number x.

- **Properties of distribution function:**

- A distribution function F(x) has the following properties:

1. $0 \leq F(x) \leq 1$ for all x.

2. $F(x)$ is nondecreasing [i.e., if $x_1 < x_2$, then $F(x_1) < F(x_2)$].

3. $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$ (since X takes on only finite values).

DISCRETE RANDOM VARIABLES

- A random variable X is said to be discrete if it can take on at most a countable number of values, say, $x_1, x_2, \ldots.$ ("Countable" means that the set of possible values can be put in a one-to-one correspondence with the set of positive integers. An example of an uncountable set is all real numbers between 0 and 1.) Thus, a random variable that takes on only a finite number of values $x_1, x_2, \ldots, x_n$ is discrete. The probability that the discrete random variable X takes on the value xi is given by

$$p(x_i) = P(X = x_i) \quad \text{for } i = 1, 2, \ldots$$

PROBABILITY MASS FUNCTION

- All probability statements about X can be computed (at least in principle) from $p(x)$, which is called the probability mass function for the discrete random variable X. The distribution function $F(x)$ for the discrete random variable X is given by

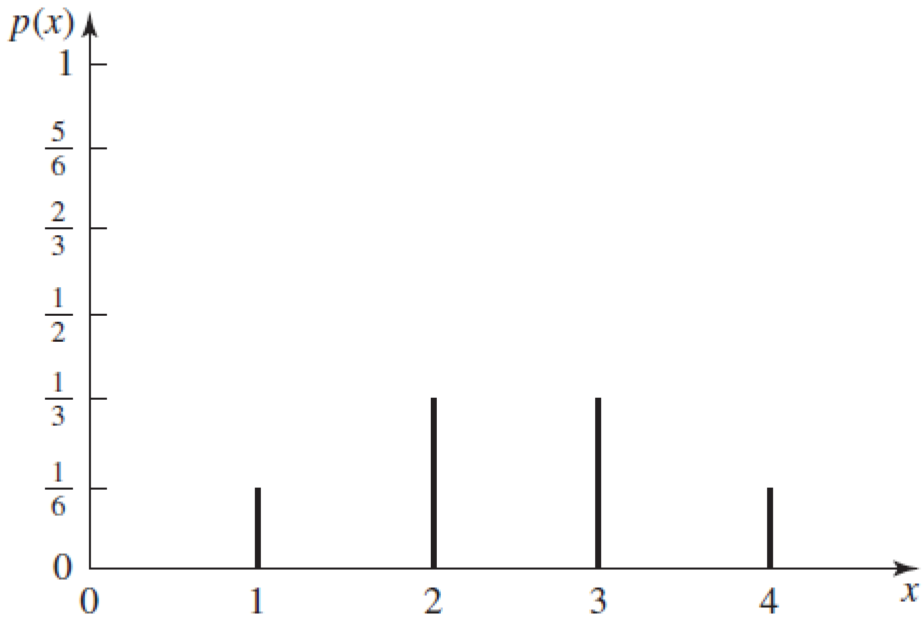
$$F(x) = \sum_{x_i < X} p(x_i) \quad \text{for all} -\infty < x < \infty$$

- **Example 5:** For the inventory example, the size of the demand for the product is a discrete random variable X that takes on the values 1, 2, 3, 4 with respective probabilities $\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6}$. The probability mass function and the distribution function for X are given in following Figs. Furthermore,

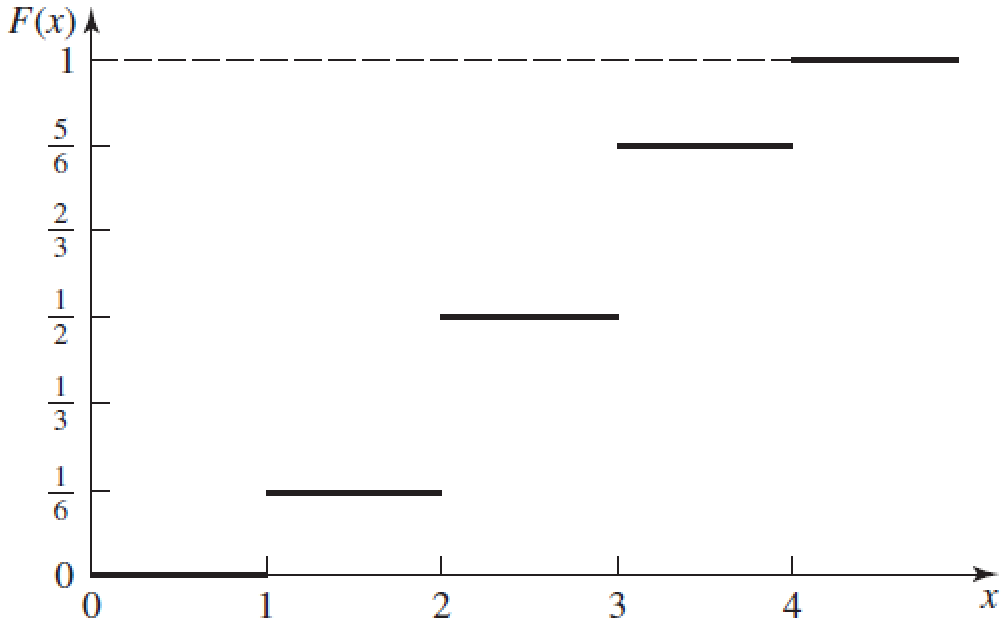
$$P(2 \leq X \leq 3) = p(2) + p(3) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

Example 6: A manufacturing system produces parts that then must be inspected for quality. Suppose that 90 percent of the inspected parts are good (denoted by 1) and 10 percent are bad and must be scrapped (denoted by 0). If X denotes the outcome of inspecting a part, then X is a discrete random variable with $p(0) = 0.1$ and $p(1) = 0.9$.

P(X) FOR THE DEMAND-SIZE RANDOM VARIABLE X



F(X) FOR THE DEMAND-SIZE RANDOM VARIABLE

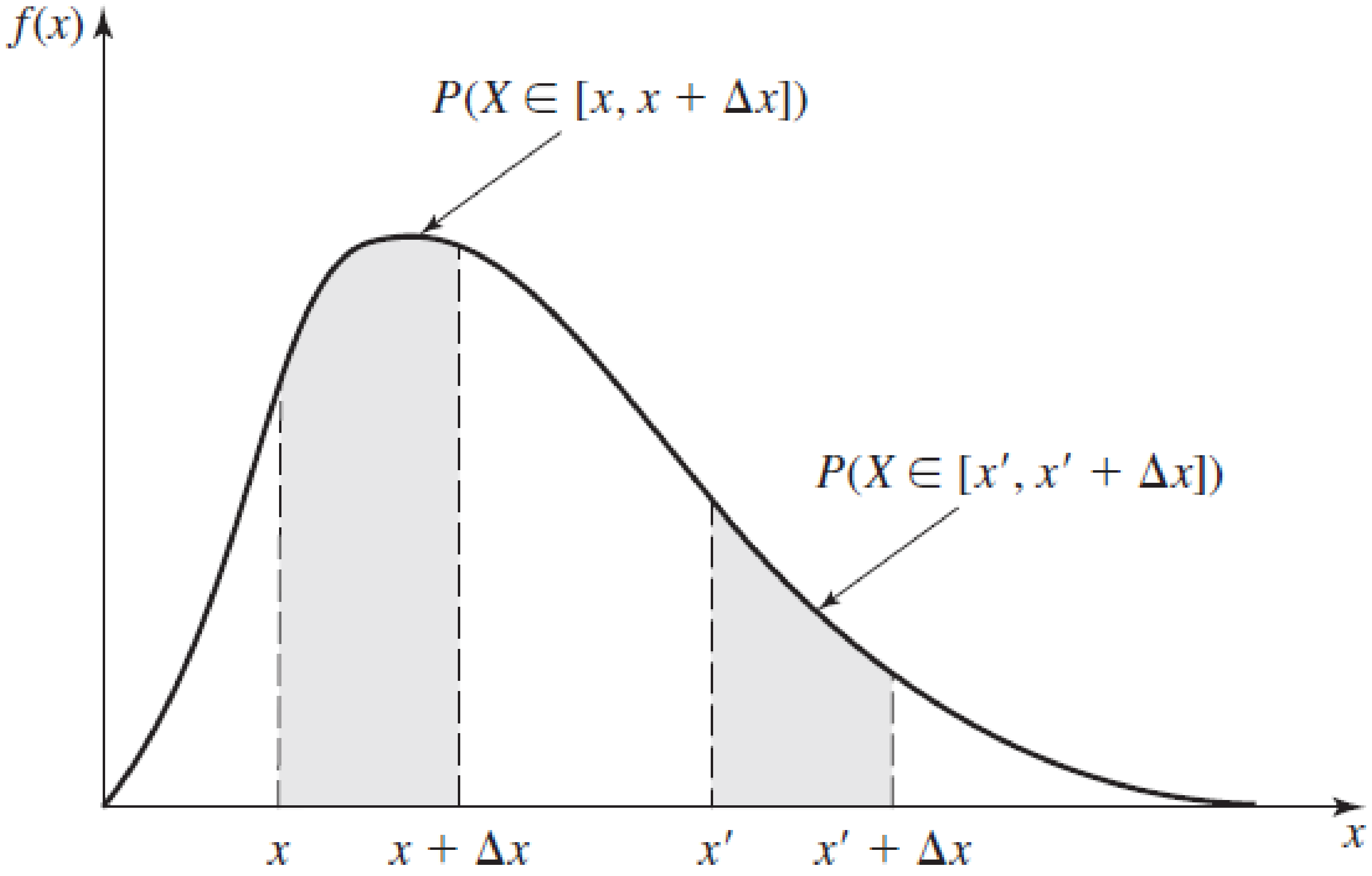


WEEK 12

SLIDES 140-145

PROBABILITY DENSITY FUNCTION

- All probability statements about X can (in principle) be computed from $f(x)$, which is called the probability density function for the continuous random variable X.



JOINT PROBABILITY MASS FUNCTION

- If X and Y are discrete random variables, then let

$$p(x,y) = P(X = x, Y = y) \quad \text{for all } x, y$$

where $p(x,y)$ is called the *joint probability mass function* of X and Y. In this case, X and Y are *independent* if

$$p(x,y) = p_X(x)p_Y(y) \quad \text{for all } x, y$$

where

$$p_X(x) = \sum_{all\ y} p(x,y)$$

$$p_Y(y) = \sum_{all\ x} p(x,y)$$

are the (marginal) *probability mass functions* of X and Y

JOINT PROBABILITY MASS FUNCTION

- **Example 7:** Suppose that X and Y are jointly discrete random variables with

$$p(x,y) = \begin{cases} \frac{xy}{27} & \text{for } x = 1,2 \text{ and } y = 2,3,4 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$p_X(x) = \sum_{y=2}^{4} \frac{xy}{27} = \frac{x}{3} \quad \text{for } x = 1,2$$

$$p_Y(y) = \sum_{x=1}^{2} \frac{xy}{27} = \frac{y}{9} \quad \text{for } y = 2,3,4$$

Since, $p(x,y) = \frac{xy}{27} = p_X(x)p_Y(y)$ for all x, y, the random variables X and Y are independent.

JOINT PROBABILITY DENSITY FUNCTION

- The random variables X and Y are jointly continuous if there exists a nonnegative function $f(x, y)$, called the *joint probability density function* of X and Y, such that for all sets of real numbers A and B,

$$P(X \in A, Y \in B) = \int_B \int_A f(x, y)\, dx\, dy$$

In this case, X and Y are *independent* if

$$f(x, y) = f_X(x) f_Y(y) \quad \text{for } all\ x, y$$

where

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

are the (marginal) *probability density functions* of X and Y, respectively

JOINT PROBABILITY DENSITY FUNCTION

- **Example 8:** Suppose that X and Y are jointly continuous random variables with

$$f(x,y) = \begin{cases} 24xy & \text{for } x \geq 0, y \geq 0, \text{and } x + y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$f_X(x) = \int_0^{1-x} 24xy\,dy \Big|_0^{1-x} = 12x(1-x)^2 \quad \text{for } 0 \leq x \leq 1$$

$$f_Y(y) = \int_0^{1-y} 24xy\,dx \Big|_0^{1-y} = 12y(1-y)^2 \quad \text{for } 0 \leq y \leq 1$$

Since

$$f\left(\frac{1}{2},\frac{1}{2}\right) = 6 \neq \left(\frac{3}{2}\right)^2 = f_X\left(\frac{1}{2}\right) f_Y\left(\frac{1}{2}\right)$$

X and Y are *not independent*.

Intuitively, the random variables X and Y (whether discrete or continuous) are *independent* if knowing the value that one random variable takes on tells us nothing about the distribution of the other. Also, if X and Y are not *independent*, we say that they are *dependent*.

WEEK 14

SLIDES 146-153

MEAN

- The *mean* is the mathematical average of a set of two or more numbers. The *mean* or expected value of the random variable X_i $(where\ i = 1, 2, \dots, n)$ will be denoted by μ_i or $E(X_i)$ and is defined by

$$\mu_i = \begin{cases} \sum_{j=1}^{\infty} x_j p_{X_i}(x_j) & \text{if } X_i \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_{X_i}(x)\, dx & \text{if } X_i \text{ is continuous} \end{cases}$$

- **Example 9:** For the demand-size random variable in inventory example, the *mean* is given by

$$\mu = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{3}\right) + 3\left(\frac{1}{3}\right) + 4\left(\frac{1}{6}\right) = \frac{5}{2}$$

MEAN

- **Example 10:** For a uniform random variable $f(x) = 1$, the *mean* is given by
- $\mu = \int_0^1 xf(x)dx = \int_0^1 x\,dx = \frac{1}{2}$
- **Properties of Mean:** Let c or c_i denote a constant (real number). Then the following are important properties of means:

1. $E(cX) = cE(X)$.
2. $E(\sum_{i=1}^{n} c_i X_i) = \sum_{i=1}^{n} c_i E(X_i)$ even if the X_i's are dependent.

MEDIAN

- The *median* is the middle number in a sorted, ascending or descending list of numbers and can be more descriptive of that data set than the average. The median $x_{0.5}$ of the random variable X_i, which is an alternative measure of central tendency, is defined to be the smallest value of x such that $F_{X_i}(x) \geq 0.5$. If X_i is a continuous random variable, then $F_{X_i}(x_{0.5}) = 0.5$.

- **Example 11:** Consider a discrete random variable X that takes on each of the values, 1, 2, 3, 4, and 5 with probability 0.2. Clearly, the *mean* and *median* of X are 3. Consider now a random variable Y that takes on each of the values 1, 2, 3, 4, and 100 with probability 0.2. The *mean* and *median* of Y are 22 and 3, respectively. Note that the *median* is insensitive to this change in the distribution.

VARIANCE

- *Variance* measures variability from the average or mean. The *variance* is a measure of the dispersion of a random variable about its *mean*. The larger the *variance*, the more likely the random variable is to take on values far from its *mean*. The variance of the random variable X_i will be denoted by σ_i^2 or $Var(X_i)$ and is defined by

$$\sigma_i^2 = E[(X_i - \mu_i)^2] = E\left(X_i^2\right) - \mu_i^2$$

- **Example 12:** For the demand-size random variable in inventory example, the *variance* is computed as follows:

$$E(X^2) = 1^2\left(\frac{1}{6}\right) + 2^2\left(\frac{1}{3}\right) + 3^2\left(\frac{1}{3}\right) + 4^2\left(\frac{1}{6}\right) = \frac{43}{6}$$

$$Var(x) = E\left(X_i^2\right) - \mu_i^2 = \frac{43}{6} - \left(\frac{5}{2}\right)^2 = \frac{11}{12}$$

- **Properties of Variance:** Let c or c_i denote a constant (real number). Then the following are important properties of variance:

1. $Var(X) \geq 0$.
2. $Var(cX) = c^2 Var(X)$.
3. $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$ if the X_i's are independent.

COVARIANCE

- *Covariance* is a measure of the relationship between two random variables. The metric evaluates how much – to what extent – the variables change together. The covariance between the random variables X_i and X_j (where $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$), which is a measure of their (linear) dependence, will be denoted by C_{ij} or $Cov(X_i, X_j)$ and is defined by

$$C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = E(X_i X_j) - \mu_i \mu_j$$

- **Example 13:** For the jointly continuous random variables X and Y in Example 8, the covariance is computed as

$$E(XY) = \int_0^1 \int_0^{1-x} xy\, f(x,y)\, dy\, dx$$

$$= \int_0^1 x^2 \left(\int_0^{1-x} 24y^2\, dy \right) dx$$

$$= \int_0^1 8x^2 (1-x)^3\, dx = \frac{2}{15}$$

$$E(X) = \int_0^1 X f_X(x)\, dx = \int_0^1 12x^2 (1-x)^2\, dx = \frac{2}{5}$$

$$E(Y) = \int_0^1 Y f_Y(y)\, dy = \int_0^1 12y^2 (1-y)^2\, dy = \frac{2}{5}$$

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) = -\frac{2}{75}$$

CORRELATION

- Correlation is a statistical measure that indicates the extent to which two or more variables fluctuate in relation to each other. A positive correlation indicates the extent to which those variables increase or decrease in parallel; a negative correlation indicates the extent to which one variable increases as the other decreases. Correlation can be determined by

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

- **Example 14:** For the random variables in example 8, it can be shown that $Var(X) = Var(Y) = \frac{1}{25}$. Therefore,

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{-\frac{2}{75}}{\frac{1}{25}} = -\frac{2}{3}$$

IMPORTANT FORMULAS TO CALCULATE CORRELATION

- Mean: $E(X) = \begin{cases} \sum_{-\infty}^{\infty} x_j p_{X_i}(x_j) & \text{if } X_i \text{ is discrete} \\ \int_{-\infty}^{\infty} x f_{X_i}(x)\, dx & \text{if } X_i \text{ is continuous} \end{cases}$
- Variance: $Var(X) = E(X_i^2) - (E(X))^2$
- Covariance: $Cov(X,Y) = E(XY) - E(X)E(Y)$
- Correlation: $Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$

WEEK 15

SLIDES 154-163

Monte Carlo Simulation

Outline

- Introduction
- History
- Examples
- Advantages
- Demonstration with Excel

What is simulation

- Simulation is the imitation of the operation of real world process or system over time.
- To engage Modelling and simulation, first create a model approximating an event.
- The model is then followed by simulation, which allows for the repeated observation of the model.
- After one or many simulations of the model, a third step takes place and that is analysis ..
- Analysis aids in the ability to draw conclusions, verify and validate the research, and make recommendations based on various iterations or simulations of the model.
- Simulation is defined to be a method that ***utilizes sequences of random numbers*** as data.

What is Monte Carlo

What is Monte Carlo Simulation?

Statistical simulation technique that provides ***approximate solution*** to ***problems*** expressed ***mathematically***.

It ***utilize*** the ***sequence of random number*** to perform the simulation.

- This techniques can be used in different domain
  - Complex Integral Computation
  - Economics Specially in Risk Management
  - extensively used in ***financial institutions*** to compute European prices,
    - to evaluate sensitivities of portfolios to various parameters and to compute risk measurements

Why Monte Carlo Simulations

- Simple implementation on computer
- Applicable for complex problems that are otherwise intractable
- Simulation does not produce an exact answer but in fact is a statistical estimate with error

The most common use of Monte Carlo Method is the evaluation of Integral and calculation of Mathematically constant variable such as PI.

History

- 1930's: ***Enrico Fermi*** uses Monte Carlo in the calculation of ***neutron diffusion***.
- *1940's*: ***Stan Ulam*** while playing ***solitaire*** tries to calculate the likelihood of winning based on the initial layout of the cards.
- After exhaustive combinatorial calculations, he ***decided*** to go for ***practical approach***
- He tries many different layouts and observing the number of successful games.
- He ***realized*** that computers could be used to solve such problems.
- ***Stan Ulam*** worked with ***John Von Neumann*** to develop ***algorithms*** including importance ***sampling and rejection sampling***.
- ***Ulam and Von Neumann*** suggested that aspects of research into ***nuclear fission*** at Los Alamos could be aided by use of computer experiments based on chance

History

- The project was ***top secret*** so ***Von Neumann*** chose the name ***Monte Carlo*** in reference to the Casino in Monaco.
- *1950's:* Many papers on Monte Carlo simulation appeared in ***physics literature***. The first major ***MCMC paper*** was published by ***Metropolis et al in 1953***.
- *1970:* Generalization of the ***Metropolis algorithm by Hastings*** which led to development of MCMC
- *1980's:* Important ***MCMC papers*** appeared in the fields of ***computer vision*** and ***artificial intelligence*** but there were few significant publications in the field of ***statistics***
- *1990:* MCMC made the first significant impact in ***statistics*** in the work of ***Gelfand and Smith.***

History

In the last 20 years MCMC has become a widely used ***tool in several fields and much research*** progress has been made.

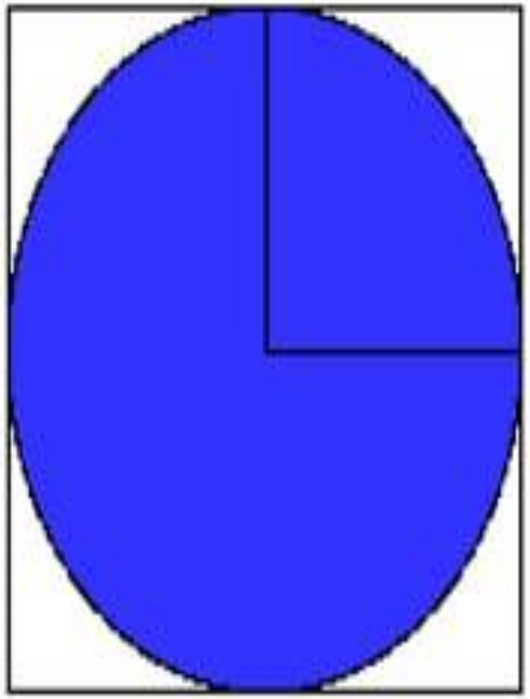
Monte Carlo Methods are now ***used to solve problems*** in numerous fields including ***applied statistics, engineering, finance and business, design and visuals, computing, telecommunications, and the physical sciences.***

WEEK 16

SLIDES 164-176

Monte Carlo Example:

Estimation of PI

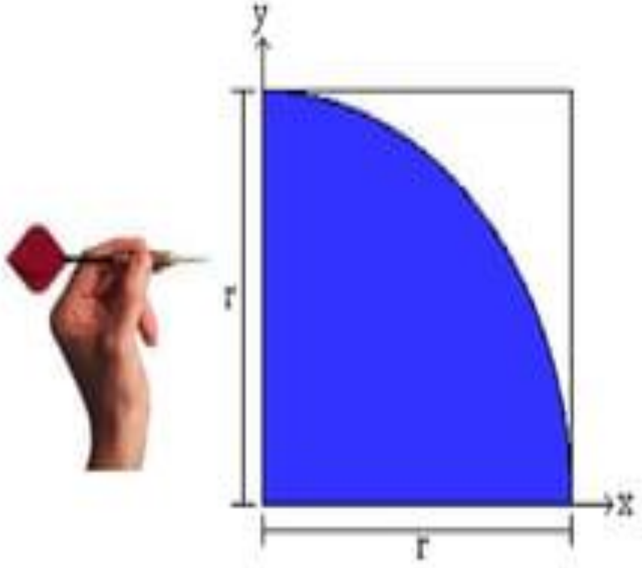


Estimating PI (Continued ..)

If you are a very poor dart player, it is easy to imagine throwing darts randomly at the figure, and

it should be apparent that of the total number of darts that hit within the square,

the number of darts that hit the shaded part (circle quadrant) is proportional to the area of that part. In other words,



$$\frac{\#\ \text{darts hitting shaded area}}{\#\ \text{darts hitting inside square}} = \frac{\text{area of shaded area}}{\text{area of square}}$$

Estimating PI (Continued ..)

If you remember your geometry, it's easy to show that

$$\frac{\#\ \text{darts hitting shaded area}}{\#\ \text{darts hitting inside square}} = \frac{\frac{1}{4}\pi r^2}{r^2} = \frac{1}{4}\pi$$

or

$$\pi = 4\,\frac{\#\ \text{darts hitting shaded area}}{\#\ \text{darts hitting inside square}}$$

Estimating PI (Continued ..)

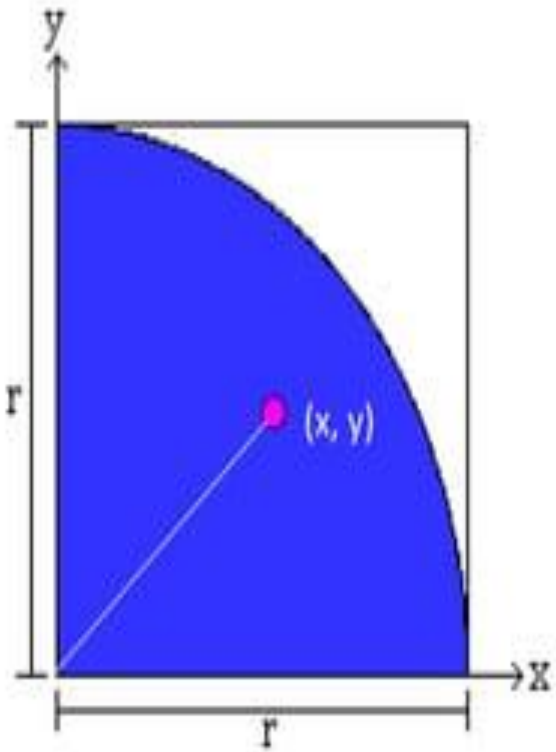
x = (random#)

y = (random#)

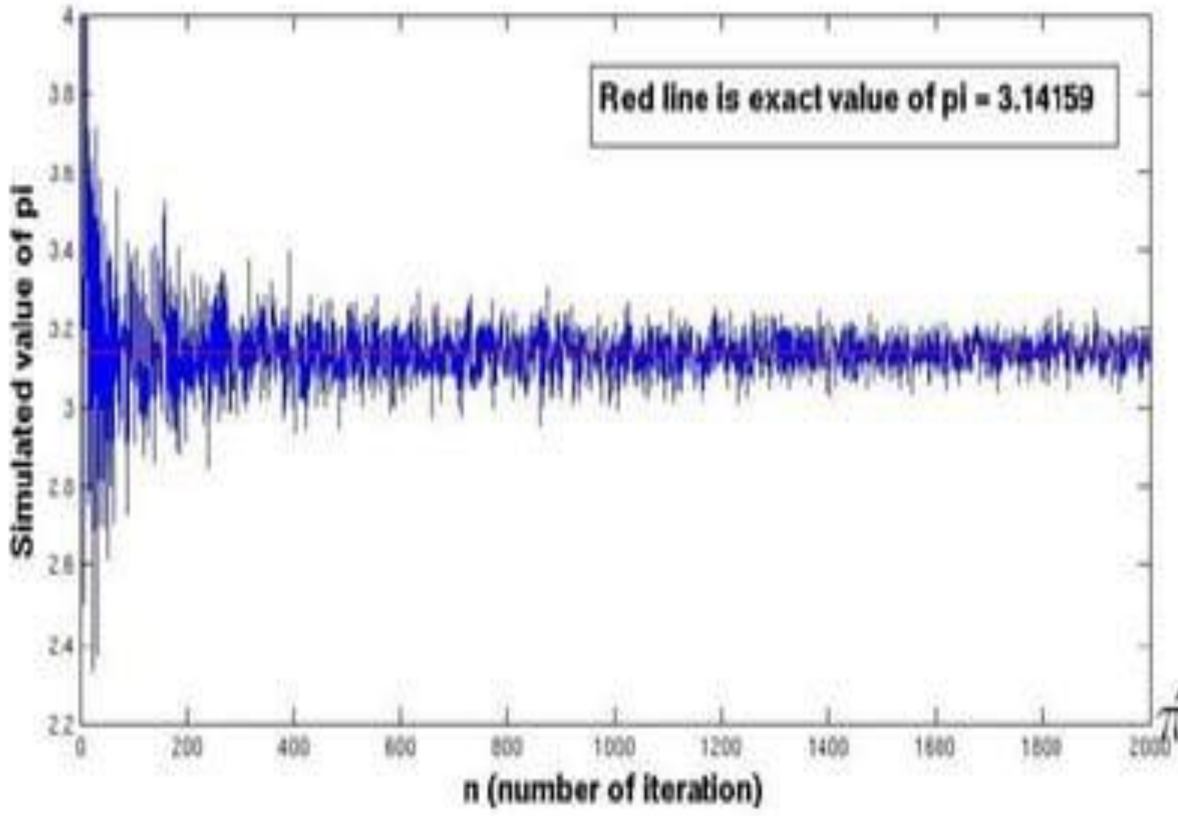
distance = sqrt (x^2 + y^2)

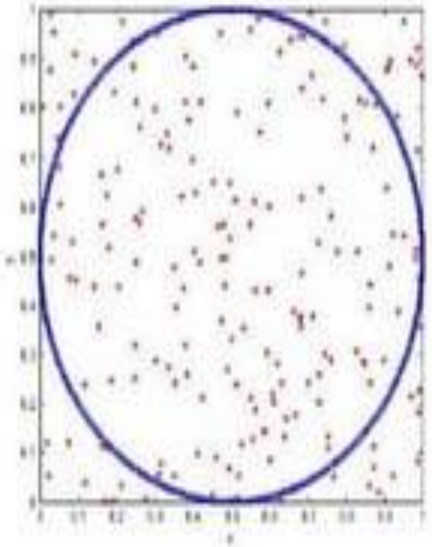
if distance.from.origin (less.than.or.equal.to) 1.0

let hits = hits + 1.0



Estimating PI (Continued ..)





$$\hat{\pi}_n = 4 \times \frac{164}{200} = 3.28.$$

A Simple Integral

Consider the simple integral:

$$I = \int_b^a f(x)\, dx$$

This can be evaluated in the same way as the *pi* example.

By randomly tossing darts at a graph of the function and tallying the ratio of hits inside and outside the function.

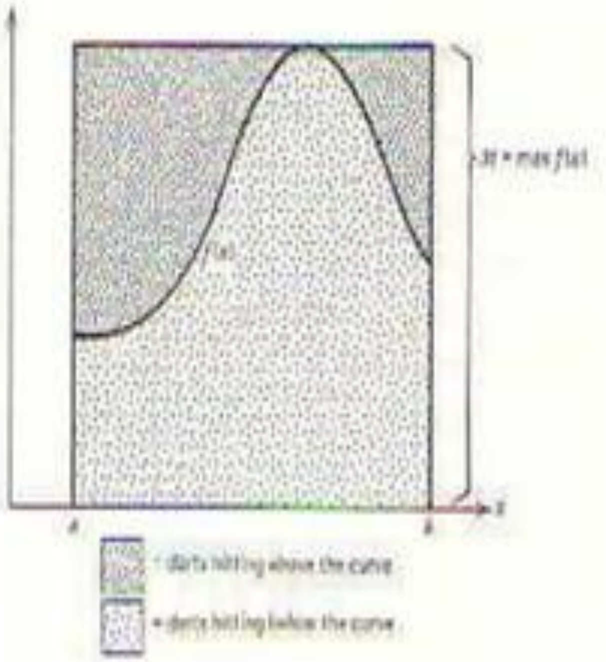
A Simple Integral (continued…)

R = {(x,y): a ≤ x ≤ b, 0 ≤ y ≤ max f(x)}

Randomly tossing 100 or so darts we could approximate the integral…

*I = [fraction under f(x)] * (area of R)*

This assumes that the dart player is throwing the darts randomly, but not so random as to miss the square altogether.



A Simple Integral (continued...)

Generally, the more iterations of the game the better the approximation will be.

- 1000 or more darts should yield a more ***accurate approximation*** of the integral than 100 or fewer.
- The ***results*** can quickly become ***skewed*** and ***completely irrelevant*** if the games ***random numbers are not sufficiently random.***

Advantages

Probabilistic Results. Results show not only what could happen, but how likely each outcome is.

Graphical Results.

- it's easy to ***create graphs of different outcomes*** and their chances of occurrence.
- This is ***important for communicating findings*** to other stakeholders.

Sensitivity Analysis.

- With just a few cases, ***deterministic analysis*** makes it ***difficult*** to see which ***variables impact the outcome*** the most.
- In Monte Carlo simulation, it's easy to see ***which inputs had the biggest effect*** on bottom-line results.

Advantages

Scenario Analysis: In deterministic models, it's very difficult to ***model different combinations*** of values for different inputs to see the ***effects of truly different scenarios***.

- Using Monte Carlo simulation, analysts can ***see exactly*** which inputs had which values together when certain outcomes occurred.

Correlation of Inputs. In Monte Carlo simulation, it's possible to ***model interdependent relationships*** between input variables.

- It's important for ***accuracy*** to represent how, in reality, when ***some factors goes up, others go up or down accordingly.***

References

1. Sabri Pllana. History of Monte Carlo method. August 2000.
   URL http://www.geocities.com/CollegePark/Quad/2435/index.html.
   http://www.geocities.com/CollegePark/Quad/2435/index.html,
2. http://www.ecs.fullerton.edu/~mathews/fofz/dirichlet/dirichle.html
3. http://mathworld.wolfram.com/DirichletProblem.html
4. http://wwitch.unl.edu/zeng/joy/mclab/mcintro.html
5. Farlow, Stanley Partial Differential Equations for Scientists and Engineers Dover Publications, New York 1982

Thank You

WEEK 17

SLIDES 177-200

CHI SQUARE TEST

CONTENTS:

- IMPORTANT TERMS
- INTRODUCTION
- CHARACTERISTICS OF THE TEST
- CHI SQUARE DISTRIBUTION
- APPLICATIONS OF CHI SQUARE TEST
- CALCULATION OF THE CHI SQUARE
- CONDITION FOR THE APPLICATION OF THE TEST
- EXAMPLE
- YATE'S CORRECTION FOR CONTINUITY
- LIMITATIONS OF THE TEST.

IMPORTANT TERMS

1) PARAMETRIC TEST: The test in which, the population constants like mean,std deviation, std error, correlation coefficient, proportion etc. and data tend to follow one assumed or established distribution such as normal, binomial, poisson etc.

2) NON PARAMETRIC TEST: the test in which no constant of a population is used. Data do not follow any specific distribution and no assumption are made in these tests. E.g. to classify good, better and best we just allocate arbitrary numbers or marks to each category.

3) HYPOTHESIS: It is a definite statement about the population parameters.

4) NULL HYPOTHESIS: (H_0) states that no association exists between the two cross-tabulated variables in the population, and therefore the variables are statistically independent. E.g. if we want to compare 2 methods method A and method B for its superiority, and if the assumption is that both methods are equally good, then this assumption is called as NULL HYPOTHESIS.

5) ALTERNATIVE HYPOTHESIS: (H_1) proposes that the two variables are related in the population. If we assume that from 2 methods, method A is superior than method B, then this assumption is called as ALTERNATIVE HYPOTHESIS.

6) DEGREE OF FREEDOM: It denotes the extent of independence (freedom) enjoyed by a given set of observed frequencies Suppose we are given a set of n observed frequencies which are subjected to k independent constraints(restrictions) then,

d.f. = (number of frequencies) – (number of independent constraints on them)

In other terms,

$$df = (r - 1)(c - 1)$$

where

r = the number of rows

c = the number of columns

7) CONTINGENCY TABLE: When the table is prepared by enumeration of qualitative data by entering the actual frequencies, and if that table represents occurance of two sets of events, that table is called the contingency table. (Latin, con- together, tangere- to touch). It is also called as an association table.

INTRODUCTION

- The chi-square test is an important test amongst the several tests of significance developed by statisticians.
- Is was developed by Karl Pearson in1900.
- CHI SQUARE TEST is a non parametric test not based on any assumption or distribution of any variable.
- This statistical test follows a specific distribution known as chi square distribution.
- In general The test we use to measure the differences between what is observed and what is expected according to an assumed hypothesis is called the chi-square test.

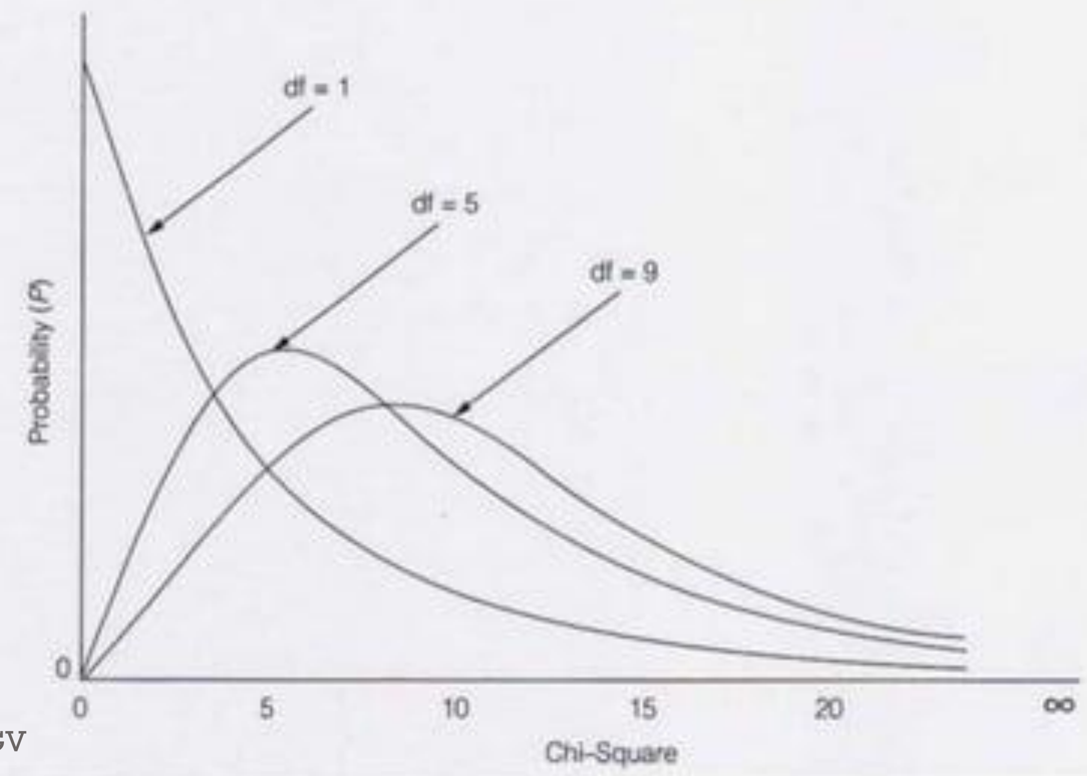
IMPORTANT CHARACTERISTICS OF A CHI SQUARE TEST

- This test (as a non-parametric test) is based on frequencies and not on the parameters like mean and standard deviation.
- The test is used for testing the hypothesis and is not useful for estimation.
- This test can also be applied to a complex contingency table with several classes and as such is a very useful test in research work.
- This test is an important non-parametric test as no rigid assumptions are necessary in regard to the type of population, no need of parameter values and relatively less mathematical details are involved.

CHI SQUARE DISTRIBUTION:

If $X_1, X_2, \ldots X_n$ are independent normal variates and each is distributed normally with mean zero and standard deviation unity, then $X_1^2+X_2^2+\ldots\ldots+X_n^2=\sum X_i^2$ is distributed as chi square (*c2*)with n degrees of freedom (d.f.) where n is large. The chi square curve for d.f. N=1,5 and 9 is as follows.

Figure 14.1 **Chi-Square Distributions for 1, 5, and 9 Degrees of Freedom**



If degree of freedom > 2 : Distribution is bell shaped

If degree of freedom = 2 : Distribution is L shaped with maximum ordinate at zero

If degree of freedom <2 (>0) : Distribution L shaped with infinite ordinate at the origin.

APPLICATIONS OF A CHI SQUARE TEST.

This test can be used in

1) Goodness of fit of distributions

2) test of independence of attributes

3) test of homogenity.

1) TEST OF GOODNESS OF FIT OF DISTRIBUTIONS:

- This test enables us to see how well does the assumed theoretical distribution (such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data.

- The $\chi 2$ test formula for goodness of fit is:

$$X^2 = \sum \frac{(o-e)^2}{e}$$

Where,
o = observed frequency
e = expected frequency

- If $\chi 2$ (calculated) > $\chi 2$ (tabulated), with (n-1) d.f, then null hypothesis is rejected otherwise accepted.

- And if null hypothesis is accepted, then it can be concluded that the given distribution follows theoretical distribution.

2) TEST OF INDEPENDENCE OF ATTRIBUTES

- Test enables us to explain whether or not two attributes are associated.
- For instance, we may be interested in knowing whether a new medicine is effective in controlling fever or not, $\chi 2$ test is useful.
- In such a situation, we proceed with the null hypothesis that the two attributes (viz., new medicine and control of fever) are independent which means that new medicine is not effective in controlling fever.
- $\chi 2$ (calculated) > $\chi 2$ (tabulated) at a certain level of significancfe for given degrees of freedom, the null hypothesis is rejected, i.e. two variables are dependent.(i.e., the new medicine is effective in controlling the fever) and if, $\chi 2$ (calculated) < $\chi 2$ (tabulated) ,the null hypothesis is accepted, i.e. 2 variables are independent.(i.e., the new medicine is not effective in controlling the fever).
- when null hypothesis is rejected, it can be concluded that there is a significant association between two attributes.

3) TEST OF HOMOGENITY

- This test can also be used to test whether the occurance of events follow uniformity or not e.g. the admission of patients in government hospital in all days of week is uniform or not can be tested with the help of chi square test.

- $\chi 2$ (calculated) < $\chi 2$ (tabulated), then null hypothesis is accepted, and it can be concluded that there is a uniformity in the occurance of the events. (uniformity in the admission of patients through out the week)

CALCULATION OF CHI SQUARE

$$X^2 = \sum \frac{(o-e)^2}{e}$$

Where,
O = observed frequency
E = expected frequency
If two distributions (observed and theoretical) are exactly alike, $\chi 2 = 0$; (but generally due to sampling errors, $\chi 2$ is not equal to zero)

STEPS INVOLVED IN CALCULATING $\chi 2$

1) Calculate the expected frequencies and the observed frequencies:

Expected frequencies f_e: the cell frequencies that would be expected in a contingency table if the two variables were statistically independent.

Observed frequencies f_o: the cell frequencies actually observed in a contingency table.

$$f_e = \frac{(\text{column total})(\text{row total})}{N}$$

To obtain the expected frequencies for any cell in any cross-tabulation in which the two variables are assumed independent, multiply the row and column totals for that cell and divide the product by the total number of cases in the table.

2) Then $\chi 2$ is calculated as follows:

$$\chi^2 = \sum \frac{(f_e - f_o)^2}{f_e}$$

CONDITIONS FOR THE APPLICATION OF $\chi 2$ TEST

The following conditions should be satisfied before X^2 test can be applied:

1) The data must be in the form of frequencies
2) The frequency data must have a precise numerical value and must be organised into categories or groups.
3) Observations recorded and used are collected on a random basis.
4) All the itmes in the sample must be independent.
5) No group should contain very few items, say less than 10. In case where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10. (Some statisticians take this number as 5, but 10 is regarded as better by most of the statisticians.)
6) The overall number of items must also be reasonably large. It should normally be at least 50.

EXAMPLE

A die is thrown 132 times with following results:

Number turned up	1	2	3	4	5	6
Frequency	16	20	25	14	29	28

Is the die unbiased?

Solution: Let us take the hypothesis that the die is unbiased. If that is so, the probability of obtaining any one of the six numbers is 1/6 and as such the expected frequency of any one number coming upward is $132 \times 1/6 = 22$. Now we can write the observed frequencies along with expected frequencies and work out the value of χ^2 as follows:

Table 10.2

No. turned up	Observed frequency O_i	Expected frequency E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$
1	16	22	–6	36	36/22
2	20	22	–2	4	4/22
3	25	22	3	9	9/22
4	14	22	–8	64	64/22
5	29	22	7	49	49/22
6	28	22	6	36	36/22

$$\therefore \qquad \sum[(O_i - E_i)^2/E_i] = 9.$$

Hence, the calculated value of $\chi^2 = 9$.

$\because$ Degrees of freedom in the given problem is

$$(n-1) = (6-1) = 5.$$

The table value' of χ^2 for 5 degrees of freedom at 5 per cent level of significance is 11.071. Comparing calculated and table values of χ^2, we find that calculated value is less than the table value and as such could have arisen due to fluctuations of sampling. The result, thus, supports the hypothesis and it can be concluded that the die is unbiased.

YATE'S CORRECTION

If in the 2*2 contingency table, the expected frequencies are small say less than 5, then $\chi 2$ test can't be used. In that case, the direct formula of the chi square test is modified and given by Yate's correction for continuity

$$\chi^2(\text{corrected}) = \frac{N \cdot \left(\left|ad - bc\right| - 0.5N\right)^2}{R1R2C1C2}$$

LIMITATIONS OF A CHI SQUARE TEST

1) The data is from a random sample.

2) This test applied in a four fould table, will not give a reliable result with one degree of freedom if the expected value in any cell is less than 5.
   in such case, Yate's correction is necessry. i.e. reduction of the mode of (o – e) by half.

3) Even if Yate's correction, the test may be misleading if any expected frequency is much below 5. in that case another appropriate test should be applied.

4) In contingency tables larger than 2*2, Yate's correction cannot be applied.

5) Interprit this test with caution if sample total or total of values in all the cells is less than 50.

6) This test tells the presence or absence of an association between the events but doesn't measure the strength of association.

7) This test doesn't indicate the cause and effect, it only tells the probability of occurance of association by chance.

8) the test is to be applied only when the individual observations of sample are independent which means that the occurrence of one individual observation (event) has no effect upon the occurrence of any other observation (event) in the sample under consideration.

THANK YOU