# STA 331 2.0 Stochastic Processes

## Introduction to Stochastic Processes

Thiyanga S. Talagala

# About

Lecturer-in-charge: Dr Thiyanga S. Talagala

Pre-requisites:

- Probability and Distribution Theory I

- Probability and Distribution Theory II,

- Programming and Data Analysis with R

# About (cont.)

Course objective:

The objective of this course is to introduce models for basic stochastic processes, in particular Markov processes.

Workload: 100 hours

# About (cont.)

**Method of Assessment:**

- Assignment/ Mid-course-unit examination: 30%

- End-course-unit examination: 70%

**Mode of contact:**

Discussion forum LMS, Online help desk, Emails

# About (cont.)

Recommended Reading:

Introduction to stochastic processes, Thiyanga S. Talagala (available online: https://thiyangt.github.io/tst.stochasticprocesses/)

Introduction to probability models. Sheldon M. Ross.

# Course content

1. Introduction to Stochastic Processes

2. Discrete Parameter Markov Chains

3. Continuous Parameter Markov Chains

# What does "stochastic" mean?

The meaning of "stochastic" is random.

# Random experiment

A **random experiment** is a physical situation whose outcome cannot be predicted with certainty until it is observed. A random experiment can be repeated as many times as we want under the same conditions (leading to different outcomes). Each one of them a trial. Thus, a trial is a particular performance of a random experiment.

# Sample space

A set of all possible outcomes of a random experiment. Here, I use $\Omega$ to denote a sample space.

*Example 1:*

*Random Experiment: Tossing of a coin.*

*Sample Space:* $\Omega = \{H, T\}$

*Example 2:*

*Random Experiment: Toss a coin three times.*

*Sample Space:*

$$\Omega = \{(H, H, H), (H, H, T), (H, T, H), (T, H, H), (H, T, T), (T, H, T),$$

$$(T, T, H), (T, T, T)\}$$

# Random variable

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A measurable mapping $X : \Omega \to \mathbb{R}$ is called a random variable.

There are two types of random variables:

i) Discrete random variable

ii) Continuous random variable

We use Roman capital letters to denote random variables ( $X$, $Y$, $Z$, $U$, $T$, etc.). However, as soon as a variable $X$ is observed, the observed values are represented by the corresponding simple Roman letter.

# Example

Consider the experiment of tossing a coin. Express the following events using a suitably defined random variable.

H = The event of getting a head

T = The event of getting a tail

# Your turn:

Let's consider a simple experiment with three possible outcomes: reporting the weather condition of a particular day as cloudy, sunny, or rainy

**Task:** Express all possible events using a suitably defined random variable.

# Your turn

suppose we monitor the weather condition every hour in a day sunny, rainy, and cloudy.

**Task:** Express all possible events using suitably defined random variable(s).

# Definition: Stochastic process

A stochastic process is a collection of random variables $\{X_t; t \in T\}$ or $\{X(t); t \in T\}$ where $T$ is an index set. That is for each $t \in T$, $X_t$ (or $X(t)$) is a random variable.

# Random variable: Probability Theory vs Stochastic Theory

**Probability theory**

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Random variable is a function $X : \Omega \to \mathbb{R}$.

**Stochastic theory**

Suppose that $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space, the function $X : T \times \Omega \to \mathbb{R}$.

We will always assume that the cardinality of $T$ is infinite, either countable or uncountable.

# Parameter space and State space

**Parameter space**

$T$ is called the parameter space.

**State space**

The set of possible values of an individual random variable $X_t$ or $X(t)$ can take is called the state space of the process.

# Classification of parameter space

**Discrete-parameter space**

When $T$ is a countable set, the process is said to be a discrete-parameter process. A discrete parameter stochastic process is denoted by $\{X_t; t \in T\}$.

**Continuous-parameter space**

When $T$ is an interval of the real line, the process is said to be a continuous-parameter process. A continuous-parameter stochastic process is denoted by $\{X(t); t \in T\}$.

# Classification of parameter space

**Discrete-parameter space**

- Observe the values of the Dow-Jones Index at the end of the $n^{th}$ week.

**Continuous-parameter space**

- The number of students waiting for a bus at any time of day.

# Classification of state space

The state space is **discrete** if it contains a finite or countably infinite number of points. Otherwise it is **continuous**.

# Discrete state space stochastic processes

**Example:** Modeling Insect Populations in a Crop Field

Let's say you're a farmer interested in understanding and managing the population of a particular pest insect in your crop field. The state space for this stochastic process is the number of insects in the field at any given time, and it's a discrete set of values (e.g., 0, 1, 2, 3, and so on).

The goal is to optimize pest control efforts to minimize crop damage while minimizing the use of pesticides.

# Discrete state space stochastic processes (cont.)

Initial State: At the beginning of the growing season, the field might be empty, so the initial state is 0 insects.

The population of insects can change over time due to various factors.

As a farmer, you can observe the insect population at certain intervals or when taking specific actions (e.g., scouting the field, applying pest control measures). These observations help you update your understanding of the current state of the insect population and adjust your pest management strategies.

# Questions

- What is the probability that the insect population in a crop field exceeds a certain threshold during the growing season, leading to potential crop damage?

- What is the expected time it takes for the pest population to reach a critical threshold, potentially leading to a pest outbreak that could harm crops?

# Questions (cont.)

- When is the optimal time to apply pest control measures to minimize crop damage and maximize crop yield?

- What is the probability that pests develop resistance to a pesticide over a given number of application cycles, and how does this affect long-term pest control strategies?

# Continuous state space stochastic processes

Air temperature at a certain place at time $t$

- When is the optimal time to plant crops, given temperature probabilities and historical data, to maximize crop yield and minimize frost risk?

- What is the probability of a heatwave, defined as a period of unusually high temperatures, lasting for a certain number of consecutive days?

Transition Probabilities: The transition probabilities represent the likelihood of temperature changes from one day to the next.

Memoryless Property: The Markov property is maintained, which means the probability of temperature on a future day depends only on the temperature on the current day and not on the past temperature history.

# Classifications of processes

1. discrete-parameter, discrete state space stochastic processes.

2. continuous-parameter, discrete state space stochastic processes.

3. discrete-parameter, continuous state space stochastic processes.

4. continuous-parameter, continuous state space stochastic processes

# Classifications of stochastic processes (cont.)

Graphs

# Question 1

Availability of a book at the time of inventory is classified as: available, misshelved, issued, missed. Suppose the inventories are conducted once every month.

i) What is the parameter space?

ii) What is the state space?

iii) What type of stochastic process is it?

# Question 2

Let $N(t)$ be the number of calls arriving at time $t$.

i) What is the parameter space?

ii) What is the state space?

iii) What type of stochastic process is it?

# Question 3

The number of customers in a queue in front of an ATM for at the end of each hour of a day. What type of stochastic process is this?

1. discrete-parameter, discrete state space stochastic processes.

2. continuous-parameter discrete state space stochastic processes.

3. discrete-parameter, continuous state space stochastic processes.

4. continuous-parameter continuous state space stochastic processes

# Question 4

Number of vehicles in parking of a shopping mall at any time during the day.

i) What is the parameter space?

ii) What is the state space?

iii) What type of stochastic process is it?

# Question 5

Classify the following stochastic process based on the state space and parameter space. A life insurance company classifies the state of health of a policy holder as Healthy, Sick, Dead. If the health status of policyholders are observed daily,

1. discrete-parameter, discrete state space stochastic processes.

2. continuous-parameter discrete state space stochastic processes.

3. discrete-parameter, continuous state space stochastic processes.

4. continuous-parameter continuous state space stochastic processes

# Question 6

Classify the following stochastic process based on the state space and parameter space. The number of particles emitted by a certain radioactive material undergoing radioactive decay during a certain period.

1. discrete-parameter, discrete state space stochastic processes.

2. continuous-parameter discrete state space stochastic processes.

3. discrete-parameter, continuous state space stochastic processes.

4. continuous-parameter continuous state space stochastic processes

# Question 7

Classify the following stochastic process based on the state space and parameter space. Daily maximum temperature observed in Colombo.

1. discrete-parameter, discrete state space stochastic processes.

2. continuous-parameter discrete state space stochastic processes.

3. discrete-parameter, continuous state space stochastic processes.

4. continuous-parameter continuous state space stochastic processes

# In this course

1. **discrete-parameter, discrete state space stochastic processes.**

2. **continuous-parameter discrete state space stochastic processes.**

3. discrete-parameter, continuous state space stochastic processes.

4. continuous-parameter continuous state space stochastic processes.

# Realization

$$X \sim Normal(4, 4)$$

```
rnorm(5, 4, 2)
```

```
[1] 0.8246619 3.9324475 5.0754297 7.3034650 4.6447904
```

```
rnorm(5, 4, 2)
```

```
[1] 6.1271962 2.1279951 1.6893880 0.5300615 3.0841193
```

Let's look at the differences

- Stochastic vs deterministic
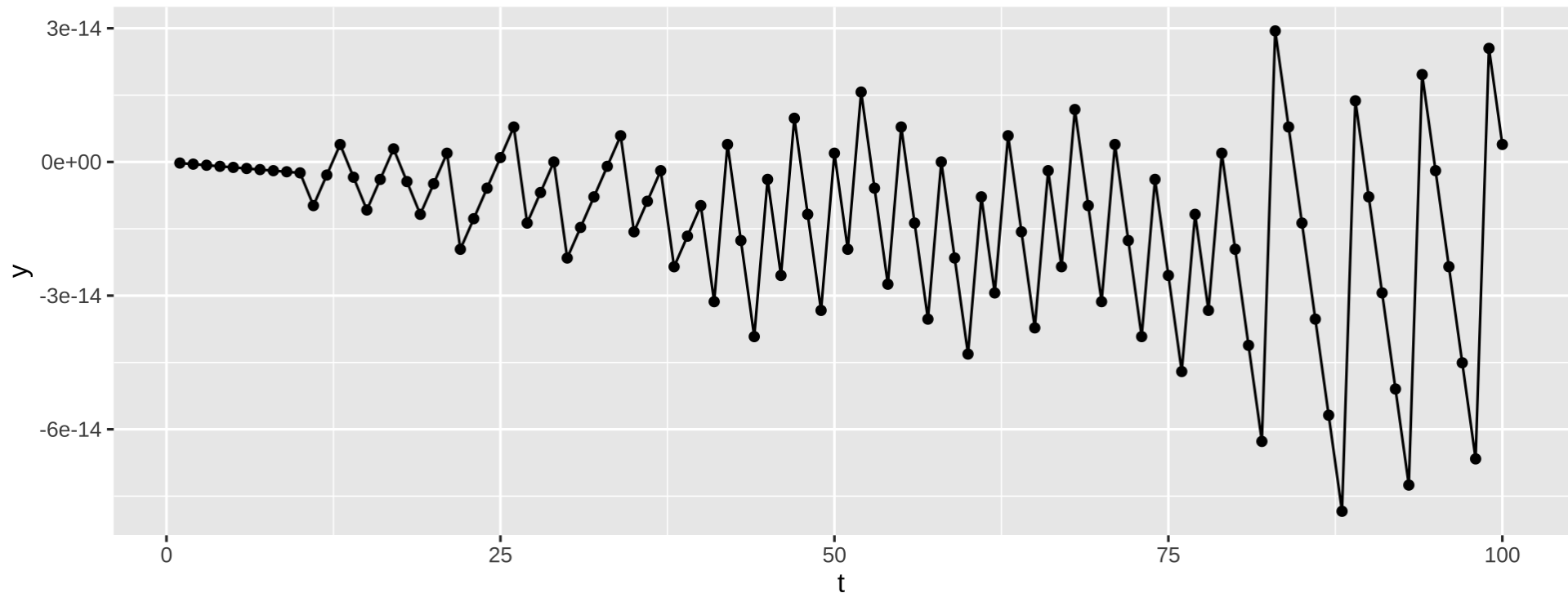
- Time series and Stochastic processes

# Stochastic process or not?

```r
t <- 1:100
y <- sin(2*pi*t)
df <- data.frame(y=y, t=t)
df
```

```
##                 y   t
## 1    -2.449294e-16   1
## 2    -4.898587e-16   2
## 3    -7.347881e-16   3
## 4    -9.797174e-16   4
## 5    -1.224647e-15   5
## 6    -1.469576e-15   6
## 7    -1.714506e-15   7
## 8    -1.959435e-15   8
## 9    -2.204364e-15   9
## 10   -2.449294e-15  10
## 11   -9.799650e-15  11
## 12   -2.939152e-15  12
## 13    3.921346e-15  13
## 14   -3.429011e-15  14
## 15   -1.077937e-14  15
## 16   -3.918870e-15  16
```
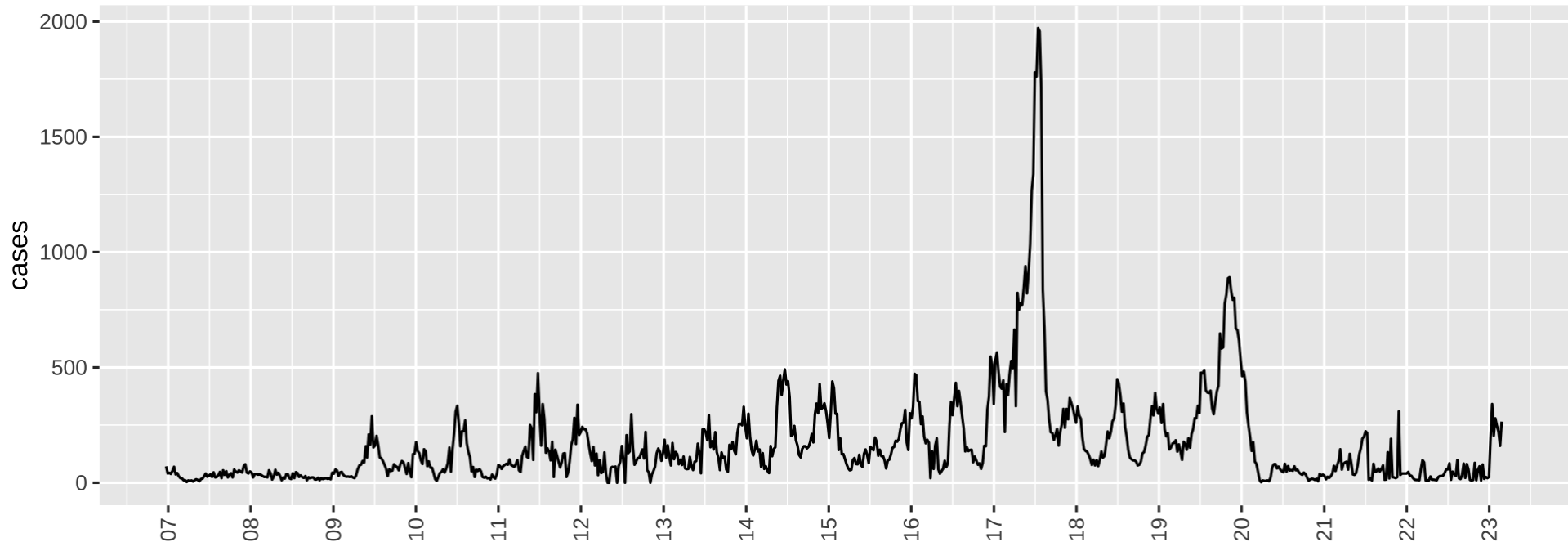
# Stochastic process or not?

```
library(ggplot2)
ggplot(data=df, aes(x=t, y=y)) +
  geom_point() +
  geom_line()
```

# Stochastic process or not?

```r
library(denguedatahub)
data(srilanka_weekly_data)
srilanka_weekly_data |>
  dplyr::filter(district == "Colombo") |>
ggplot(aes(x=start.date, y=cases)) + geom_line()  + scale_x_date(date_breaks = "1 year",
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

# Applications

- Population dynamics

- Genome evolution

- Statistical pattern of arrivals/ waiting-line analysis or queueing problem of operations research

- Service mechanism describes when service is available, how many customers can be served at a time, and how long service takes

- Applications to risk theory, insurance, actuarial science and system risk engineering

# Example 1

Suppose that a virus can exist in 4 different strains (species), numbered from 1 to 4. In each new generation it either stays the same, or with probability $\alpha$ mutates to another strain, chosen at random. The $\alpha$ is $0 < \alpha < 1$.

Explain how to compute the probability that the strain in the fifth generation is the same as the initial strain?

# Example 2

The weather changes at a tourist resort from one day to the next can somewhat simplified be described as a Markov chain with the three states: Sunny, Cloudy and Rainy. Using the weather statistics of the area the following transition probability matrix has been estimated.

$$P = \begin{bmatrix} & S & C & R \\ S & 0.5 & 0.3 & 0.2 \\ C & 0.3 & 0.5 & 0.2 \\ R & 0.6 & 0.1 & 0.3 \end{bmatrix}$$

A tourist intends to visit the resort during 24 - 26 October. What is the probability that there will be three sunny days in a row?

# Example 3

Individuals arrive to a COVID-19 vaccination centre according to a nonhomogeneous Poisson process having the rate function

$$\lambda(t) = \begin{cases} 2t & \text{for } 0 \leq t < 1, \\ 2 & \text{for } 1 \leq t < 2, \\ 4 - t & \text{for } 2 \leq t < 4, \end{cases}$$

where $t$ is measured in hours from 8.00am.

Calculate the probability that two people arrive from 8.00am to 10.00am?