Lesson 4. Introduction to Stochastic Processes and the Poisson Arrival Process

1 What is a stochastic process?

- A **stochastic process** is a mathematical model of a probabilistic system that evolves over time and generates a sequence of numerical values
  - Each numerical value in the sequence is modeled by a random variable
- Examples:
  - the sequence of daily prices of a stock
  - the sequence of failure times of a machine
  - the sequence of radar measurements of the position of an airplane
- Another perspective: a stochastic process is a sequence of random variables ordered by an index set
- Examples:
  - $\{S_n; n = 0, 1, 2, \dots\} = \{S_0, S_1, S_2, \dots\}$ with discrete index set $\{0, 1, 2, \dots\}$
  - $\{Y_t; t \geq 0\}$ with continuous index set $\{t \geq 0\}$
- The indices n and t are often referred to as "time"
  - $\{S_n; n = 0, 1, 2, \dots\}$ is a **discrete-time process**
  - $\{Y_t; t \geq 0\}$ is a **continuous-time process**
- The **state space** of a stochastic process is the range (possible values) of its random variables
  - State spaces can be discrete or continuous
    (i.e. the random variables of a stochastic process can be discrete or continuous)

2 The Case of the Reckless Beehunter

Citizens of Beehunter have complained that a busy intersection has recently become more dangerous, and they are demanding that the city council take action to make the intersection safer. The city council agrees to undertake a study of the intersection to determine if the accident rate has actually increased above the 1 per week average that is (unfortunately) considered normal. It hires a traffic engineer from nearby Vincennes, to perform the study.

The traffic engineer recommends that the number of accidents at the intersection be recorded for a 24-week period. Based on historical data, she has determined that the time between accidents is exponentially distributed with a mean of 1 week. If the number of accidents is significantly larger than expected, then she will declare that the intersection has indeed become more dangerous. During the study period, 36 accidents were observed.

- Our approach:
  - Model the system as a stochastic process
    - ◇ Time between accidents as random variables
    - ⇒ Time of the nth accident, number of accidents by time t
  - Analyze the model to determine if the probability that 36 accidents occur in a 24-week period is "small"

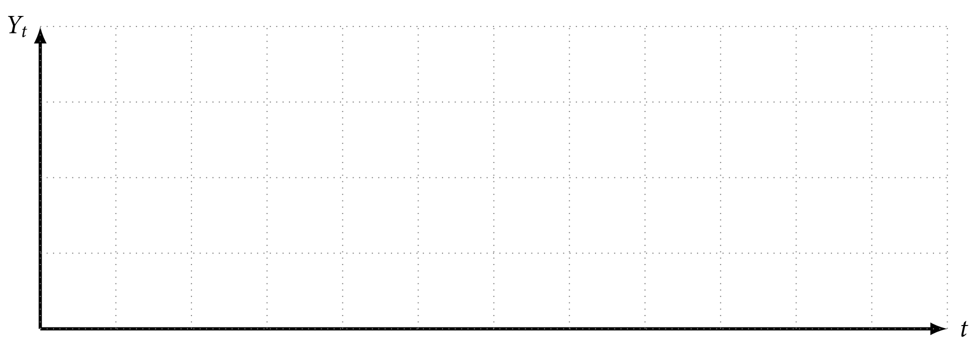
3 Arrival counting processes

- Suppose we want to count the number of "arrivals" to a system
- An "arrival" is broadly defined as any discrete unit that can be counted: for example,
  - customer arrivals
  - service requests
  - accidents at an intersection
- Let

$$\begin{aligned} G_n &= \text{the } n\text{th } \textbf{interarrival time}\text{, or the time between the } (n-1)\text{th arrival and the } n\text{th arrival} \\ T_n &= \text{the } n\text{th } \textbf{arrival time}\text{, or the time of the } n\text{th arrival} \\ Y_t &= \text{the number of arrivals by time } t \end{aligned}$$

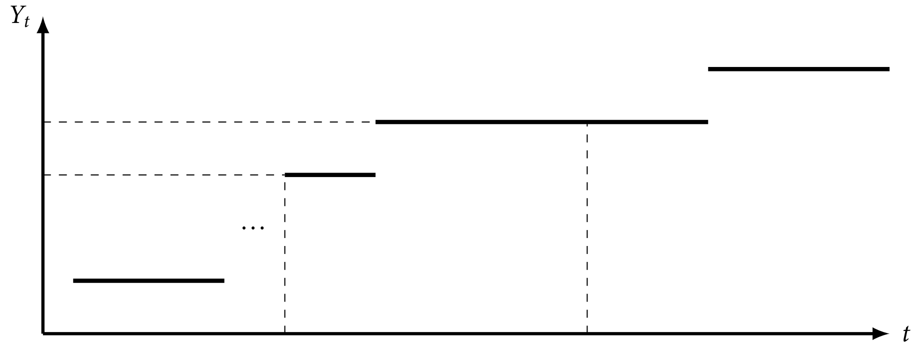
- We assume the system is empty at time 0:
- We model the interarrival times as random variables
  - ⇒ The arrival times and number of arrivals are also random variables
  - ⇒ $\{Y_t; t \geq 0\}$ is a stochastic process!
- This type of stochastic process is known as a **arrival counting process**

Example 1. Suppose the random variates of G_1, G_2, and G_3 are 3, 2, and 4, respectively. Graph Y_t.



- We can write T_n in terms of the interarrival times $G_1, \ldots, G_n$:

- The number of arrivals Y_t and the arrival times T_n are fundamentally related:



4 The Poisson arrival process

- Now, let's assume that the interarrival times $G_1, G_2, \ldots$ are independent exponential random variables with common parameter λ:

$$F_{G_n}(a) = 1 - e^{-\lambda a} \quad \text{for } a \geq 0 \qquad E[G_n] = \frac{1}{\lambda} \qquad \text{Var}(G_n) = \frac{1}{\lambda^2}$$

  - Note that the interarrival times are **time stationary**: the distribution stays the same over time

- This kind of arrival counting process is known as a **Poisson arrival process with arrival rate λ**
  - Why "Poisson" and "arrival rate"? We'll see soon.

- Since $G_1, \ldots, G_n$ are independent exponential random variables with parameter λ, $T_n = G_1 + \cdots + G_n$ is an **Erlang random variable with parameter λ and n phases**:

$$F_{T_n}(a) = 1 - \sum_{j=0}^{n-1} \frac{e^{-\lambda a}(\lambda a)^j}{j!} \quad \text{for } a \geq 0 \qquad E[T_n] = \frac{n}{\lambda} \qquad \text{Var}(T_n) = \frac{n}{\lambda^2}$$

- Therefore, we can get an explicit expression for the cdf of Y_t:

- And we can also get an expression for the pmf of Y_t:

- The pmf and cdf may look familiar: Y_t is a **Poisson random variable** with parameter λt

$$\Rightarrow \quad E[Y_t] = \lambda t \qquad \text{Var}(Y_t) = \lambda t$$

- The average arrival rate in this arrival process is

> **Example 2.** In the Beehunter case, the inter-accident times were exponentially distributed with parameter $\lambda = 1$. What is the probability that the total number of accidents at week 24 is greater than 36? What is the expected number of accidents? (Your calculator can compute summations.)

5 Properties of the Poisson process

- Let $\Delta t > 0$ be a time increment
- The **forward-recurrence time** R_t is the time between t and the next arrival
- The **independent-increments property**: the number of arrivals in nonoverlapping time intervals are independent random variables:
  - As a consequence:
- The **stationary-increments property**: the number of arrivals in a time increment of length Δt only depends on the length of the increment, not when it starts:
  - As a consequence:

  $\Rightarrow$ λ can be interpreted as the **arrival rate** of the Poisson process
- The **memoryless property**: the forward-recurrence time R_t has the same distribution as the interarrival time:
- These properties make computing probability statements about Poisson processes pretty easy

Example 3. Recall that in the Beehunter case, a total of 103 accidents have occurred at the intersection up to the time the traffic engineer starts observing, time a. What is the probability that more than 36 accidents are observed in the following 24 weeks? What is the expected number of accidents?

Example 4. What is the probability there are 4 accidents at week 5, given that there are 2 accidents at week 4? What is the expected number of accidents at week 5?

Example 5. What is the probability that the 10th arrival occurs before the 7th week? What is the expected time of the 10th arrival?

6 When is the Poisson process a good model?

- **Any arrival-counting process in which arrivals occur one-at-a-time and has independent and stationary increments must be a Poisson process**
  - If you can justify your arrivals having independent and stationary increments, then you can assume that the interarrival times are exponentially distributed
  - This is a very powerful result
- Independent increments $\Leftrightarrow$ number of arrivals in nonoverlapping intervals of time are independent
  - Reasonable when the arrival-counting process is formed by a large number of customers making individual, independent decisions about when to arrive
- Stationary increments $\Leftrightarrow$ expected number of arrivals = constant rate $\times$ length of time interval
  - Reasonable when arrival rate is approximately constant over time

Example 6. Discuss whether or not it is reasonable to approximate the following arrival processes as Poisson processes:

a. The arrival of cars at a toll booth during evening rush hour.
b. The arrival of students at a college football game.

7 Why does the memoryless property hold?

- The memoryless property allows us to ignore when we start observing the Poisson process, since forward-recurrence times and interarrival times are distributed in the same way
- "Memoryless" $\longleftrightarrow$ how much time has passed doesn't matter
- Why is this true for Poisson processes?
- Let's consider G_n, the interarrival time between the $(n-1)$th and nth arrival (between T_{n-1} and T_n)
  - Recall that $G_n \sim \text{Exponential}(\lambda)$
- Pick some t between T_{n-1} and T_n
- We want to show that the forward-recurrence time $R_t \sim \text{Exponential}(\lambda)$
  - Equivalently, we show $F_{R_t}(a) = \Pr\{R_t \leq a\} = 1 - e^{-\lambda a}$

- Therefore:

$$\begin{aligned}
\Pr\{R_t > a\} &= \Pr\{G_n > t - T_{n-1} + a \,|\, G_n > t - T_{n-1}\} \\
&= \frac{\Pr\{G_n > t - T_{n-1} + a \text{ and } G_n > t - T_{n-1}\}}{\Pr\{G_n > t - T_{n-1}\}} \\
&= \frac{\Pr\{G_n > t - T_{n-1} + a\}}{\Pr\{G_n > t - T_{n-1}\}} \\
&= \frac{e^{-\lambda(t - T_{n-1} + a)}}{e^{-\lambda(t - T_{n-1})}} = e^{-\lambda a}
\end{aligned}$$

$$\Rightarrow \quad \Pr\{R_t \leq a\} = 1 - e^{-\lambda a}$$

- Note: This "proof" is rough and sketchy – we actually need to condition on T_{n-1} and Y_t
  - Repeated use of the law of total probability
- The independent-increments and stationary-increments properties follow from the memoryless property and the fundamental relationship between Y_t and T_n (see Nelson pp. 110-111)

8 Exercises

Problem 1. The Markov Company has a manufacturing cell that processes jobs during a 12-hour shift starting at 6 a.m. and ending at 6 p.m. Jobs leave the cell according to a Poisson process with rate $\lambda = 8$ per hour.

a. If the cell has processed exactly 10 jobs by 8 a.m., what is the probability that the cell will have processed more than 30 jobs by 10 a.m.?
b. What is the probability that the cell will have processed its 50th job before 12 p.m.?
c. If the cell has processed exactly 40 jobs by 12 p.m., what is the probability that the cell will have processed its 100th job by the end of the shift?
d. When are the first 4 jobs expected to be completed? (Assume all jobs are available starting at 6 a.m.)