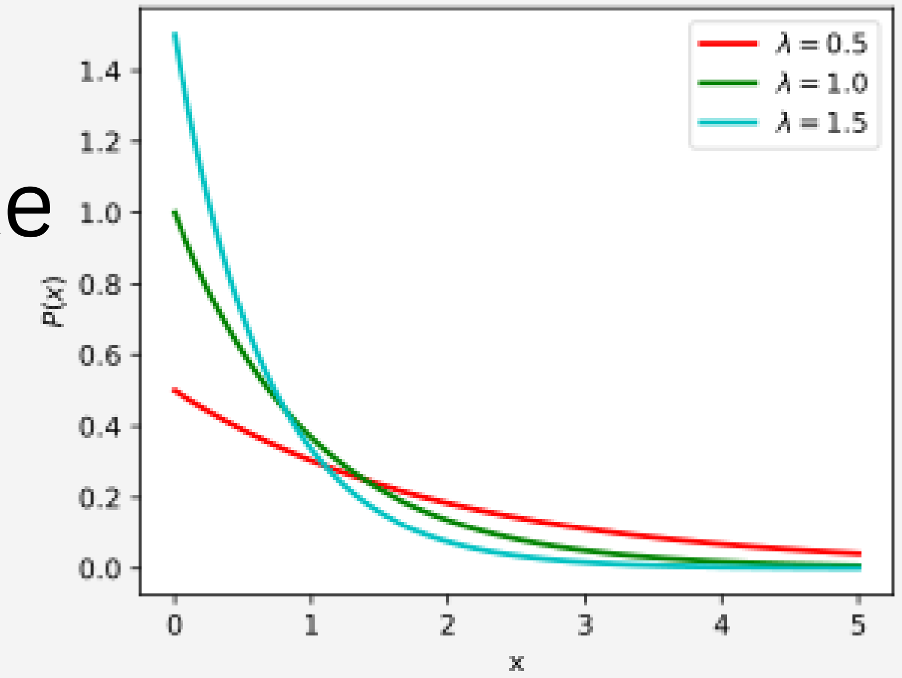
Sep 18, 2024

CS 350 DISCUSSION 3

Probabilities in System Analysis

Probabilities review

Exponential distribution: rate parameter λ representing the rate of event occurrence



$\boxed{CDF}$

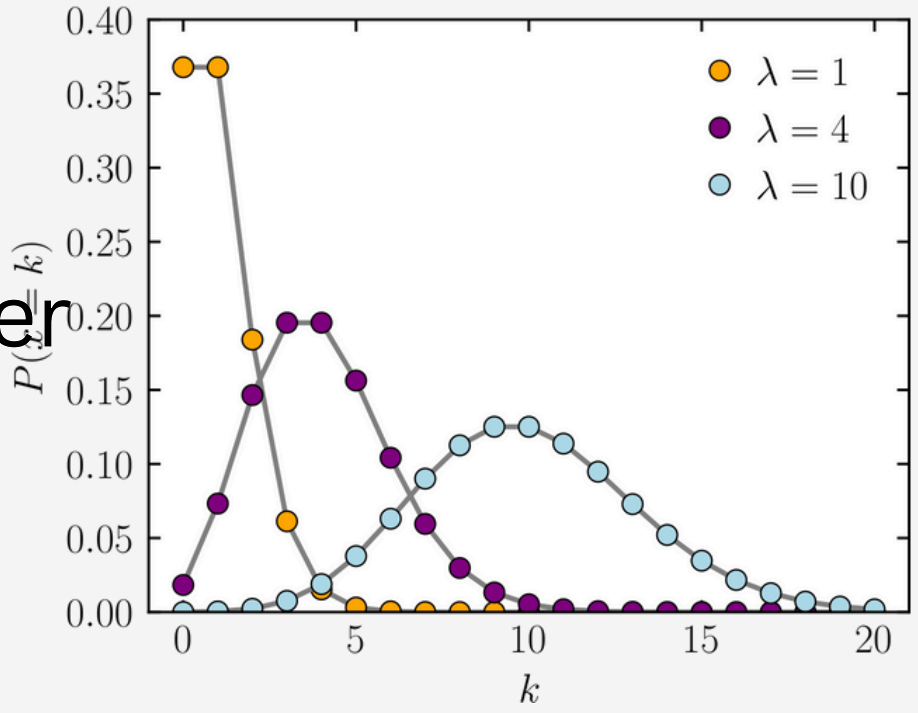
$$P(T \le t) = 1 - P(T > t) = 1 - e^{-\lambda t}$$

$$\Downarrow$$

$\boxed{PDF}$

$$\frac{d}{dt}(CDF) = \frac{d}{dt}(1 - e^{-\lambda t}) = \lambda e^{-\lambda t}$$

Poisson distribution is a discrete distribution that represents the number of events that occur in an interval of time, given the mean arrival rate λ



PMF	$\frac{\lambda^k e^{-\lambda}}{k!}$
CDF	$e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$

Let's Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

a) What is the throughput of the CodeBuddy system?

Problem 5.2 from the book

Let's Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

a) What is the throughput of the CodeBuddy system?

The arrival rate of the server is also the rate of departure, which is the throughput of the server at steady state. We know that λ=1/inter-arrival time

1/45=0.022 req/s

Problem 5.2 from the book

Let's Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

b) What is the formula for the probability distribution of the inter-arrival time of requests to the CodeBuddy system?

Problem 5.2 from the book

Let’s Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

b) What is the formula for the probability distribution of the inter-arrival time of requests to the CodeBuddy system?

We assume the inter-arrival times follow exponential distribution.
The PDF is of an exponentially distributed function is: PDF = $f(x) = \lambda e^{(-\lambda x)}$

Substituting the value of λ, we have $0.022e^{(-0.022x)}$

Problem 5.2 from the book

Let’s Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

c) What is the probability that two subsequent submissions will be separated from one another by more than 1 minute?

Problem 5.2 from the book

Let's Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

c) What is the probability that two subsequent submissions will be separated from one another by more than 1 minute?

$$P(t > 60s) = 1 - P(t \leq 60s) = 1 - F(60) = 1 - (1 - e^{-0.022 \times 60}) = e^{-1.32} = 0.27$$

Problem 5.2 from the book

Let’s Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

d) What is the standard deviation of the inter-arrival time of requests to the CodeBuddy system?

Problem 5.2 from the book

Let’s Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

d) What is the standard deviation of the inter-arrival time of requests to the CodeBuddy system?

Standard deviation of exponential distributions = their mean/expectation = $1/\lambda$

$1/0.022=45.5$

Problem 5.2 from the book

Let's Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

e) What is the formula for the probability distribution of the number of requests arriving to the CodeBuddy system in a period of 120 seconds.

Problem 5.2 from the book

Let's Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

e) What is the formula for the probability distribution of the number of requests arriving to the CodeBuddy system in a period of 120 seconds.

Since the arrival rate follows exponential distribution, the number of requests arriving within an interval of time follows Poisson distribution.

$$\text{PMF} = f(x) = \frac{(0.022 \times 120)^x e^{-0.022 \times 120}}{x!} = \frac{(2.64)^x e^{-2.64}}{x!}$$

Problem 5.2 from the book

Let's Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

f) What is the probability that more than 3 submissions will be received within a 120-second period of time?

Problem 5.2 from the book

Let’s Practice

Exercise 1

Take a look at how CodeBuddy works. New submissions are queued up in their order of arrival and then processed one at a time. Fast-forward to a future when Renato has implemented a feature that displays the average inter-arrival time between student submissions and that such value is 45 seconds. Figure out the following about the performance of CodeBuddy.

f) What is the probability that more than 3 submissions will be received within a 120-second period of time?

Per the previous question’s PMF:

$$
\begin{aligned}
P(x > 3) &= 1 - P(x \leq 3) \\
&= 1 - [P(x = 3) + P(x = 2) + P(x = 1) + P(x = 0)] \\
&= 1 - 0.728 \\
&= 0.272
\end{aligned}
$$

Problem 5.2 from the book

Let’s Practice

Exercise 2

You have just been promoted to the role of Quality of Service chief at the RelEyeAnt Corp. which provisions computing systems with reliability guarantees. The subsystem you are currently working is comprised of three machines, namely S1, S2, and S3. Requests arriving at the subsystem are of two types: simple and complex. Simple requests require service only at one of these machines, and it is up to you to decide which machine should serve which request. A simple request will take the same time to execute on any of these machines. Complex requests on the other hand need to be served at S1 first, then they move to S2 and then to S3. After having conducted a long analysis on the three machines, you have been reported that S1 has an availability of 98% and an MTBF of 1 hour; S2 has an availability of 91% and a MTBF of 10 hours; and S3 has an availability of 85% and a MTBF of 100 hours. Figure out how to satisfy the customer contracts described below. Always motivate your reasoning and recall that reliability is defined as $R(t) = e^{-t/MTBF}$*. A simple request (or a chunk of a complex request) is correctly processed by a machine if it arrives at a machine that is correctly operating, and if that machine remaining in operational state until the request completes processing*

a) Customer A wants to submit simple requests that require 6 hours of processing. She also demands that a submitted request should be correctly processed with probability 70%. To which machine shall we route her requests to to honor the deal?

Problem 2 from Midterm 1, Fall 2021

Let's Practice

Exercise 2

a) Customer A wants to submit simple requests that require 6 hours of processing. She also demands that a submitted request should be correctly processed with probability 70%. To which machine shall we route her requests to to honor the deal?

We know that the availability of our servers is $\alpha_1 = 0.98$, $\alpha_2 = 0.91$, and $\alpha_3 = 0.85$.

Handling Customer's A requests at S1 would yield a probability of correct completion of $p_{serv} = \alpha_1 \cdot R_1(6) = 0.0024$.

Handling Customer's A requests at S2 would yield a probability of correct completion of $p_{serv} = \alpha_2 \cdot R_2(6) = 0.5$.

Handling Customer's A requests at S3 would yield a probability of correct completion of $p_{serv} = \alpha_3 \cdot R_3(6) = 0.8$.

So our only option is to handle her requests at S3.

Problem 2 from Midterm 1, Fall 2021

Let's Practice

Exercise 2

b) Customer B wants to submit complex requests that would take 15 minutes to complete at S1, 30 minutes at S2, and 20 hours at S3. What is the probability that a generic request submitted by this customer will be correctly served?

Let's compute $R_1(0.25) = e^{-0.25/1} = 0.779$, $R_2(0.5) = e^{-0.5/10} = 0.95$, $R_3(20) = e^{-20/100} = 0.819$.

Now the probability that a request is processed correctly means that each chunk arrives at the corresponding machine while the machine is up, and completes without the machine going offline mid-request. Therefore:

$$p_{serv} = \alpha_1 \cdot R_1(0.25) \cdot \alpha_2 \cdot R_2(0.5) \cdot \alpha_3 \cdot R_3(20) = 0.459$$

Problem 2 from Midterm 1, Fall 2021

Let’s Practice

Exercise 2

c) Consider the following approach for fault tolerance. Simple requests are always first routed to S1. If they fail there, they are routed to S2. If they fail there, they are routed to S3. Finally, if the also fail at S3, they are aborted. Out of a sequence of identical requests that demand 2 hours of processing, how many on average will be handled using S1 or S2, before one needs to be sent to S3? Note: the one which uses S3 shall be excluded from the average.

Let’s compute the probability of “failure” in the Bernoulli trial that consists in submitting a request to the system. In this case is the probability of failure corresponds to that of being correctly handled at either S1 or S2.

$$p_{fail} = \alpha_1 \cdot R_1(2) + (1 - \alpha_1 \cdot R_1(2)) \cdot \alpha_2 \cdot R_2(2) = 0.133 + 0.646 = 0.78$$

The number of identical 2-hours requests that will be handled by S1 or S2 without involving S3 follows a Geometric distribution. Therefore the average is: $\mu_{S1 or S2} = \frac{p_{fail}}{1-p_{fail}} = 3.5$

Problem 2 from Midterm 1, Fall 2021