Poisson Distribution

Instructor, Nero Chan Zhen Yu

Forward School

Checkout Line

- Let's say that you are a cashier at Tesco. It is 6.30pm and your shift ends at 7.00pm. The store policy is to close your checkout line 15 minutes your shift ends (in this case 6.45pm) so you can finish checking out the customer already in your line and leave on time.

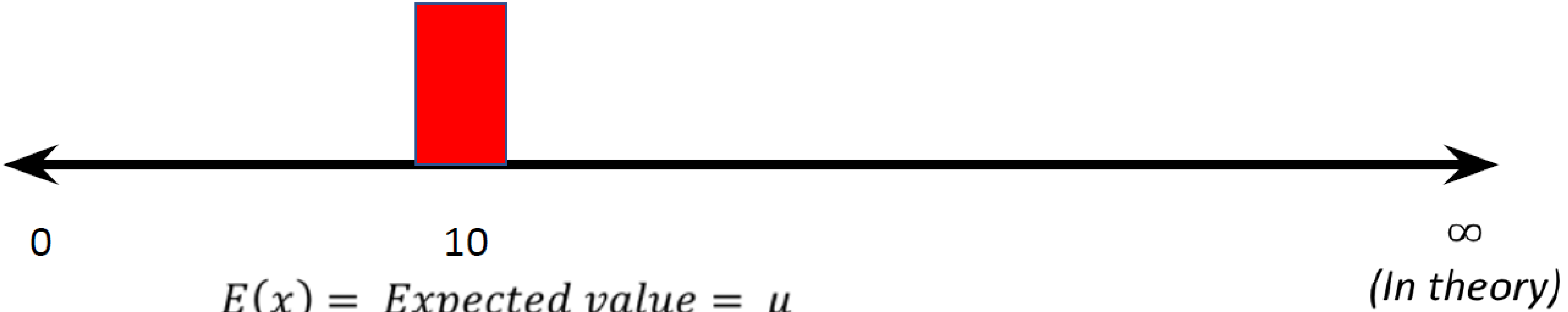
- By examining the CCTV, store data indicates that between 6.30pm and 6.45pm each weekday, an average of 10 customers enter any given checkout line.

- Questions:
  - What is the probability that exactly 7 customers enter your line between 6.30pm to 6.45pm.
  - What is the probability that more than 10 customers arrive (which mean you will probably need to work later after your shift, later than 7.00pm)

Forward School

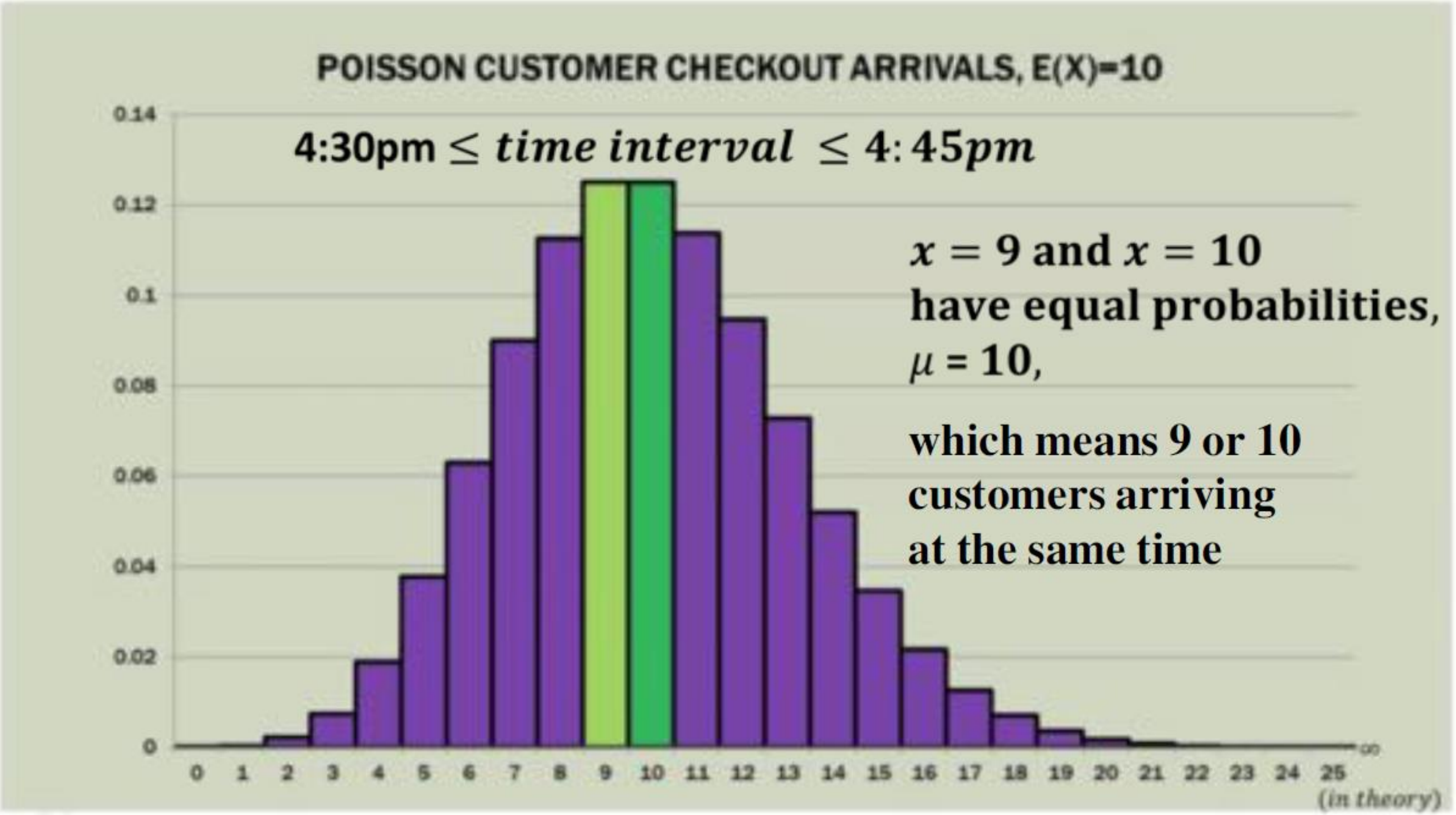
Checkout Line

- Between 6:30 pm – 6:59:59pm
- Which outcome (#customer) is your expected value/mean?



$$E(x) = \text{Expected value} = \mu$$

Checkout Line



Forward School

Poisson Distribution

- Poisson distribution focuses on the number of discrete events or occurrences over a specified interval or continuum (time, length, distance etc)

$$\lambda = \frac{\#\ occurrences}{specified\ intervals}$$

"lambda"

$$E(x) = Expected\ value = \mu = \lambda$$

Checkout line

$$\lambda = \frac{10\ customers}{per\ 15\ minutes} = 10$$

Forward School

Poisson Characteristics

1. Discrete outcomes ($x = 0, 1, 2, 3\ \ldots$)
2. The number of occurrences in each interval can range from zero to infinity (theoretically); $0 \leq x\ \leq \infty$
3. Describes the distribution of infrequent (rare) events
4. Each event is independent of the other events
5. Describes discrete events over an interval (time, distance, etc)
6. Expected number of occurrences $E\ (X)$ are assumed to be constant throughout the experiment

Forward School

Poisson Formula

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ or } P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

$x = 0, 1, 2, 3 \ \ldots \infty$; # occurrences of interest

λ or μ = long run average = $\frac{\#\ occurrences}{interval}$

e = 2.718282 (base of natural logs)

Forward School

Probability of exactly 7 customers

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

What is the probability that exactly 7 customers enter your line between 6:30pm – 6:45pm?

$x = 7$

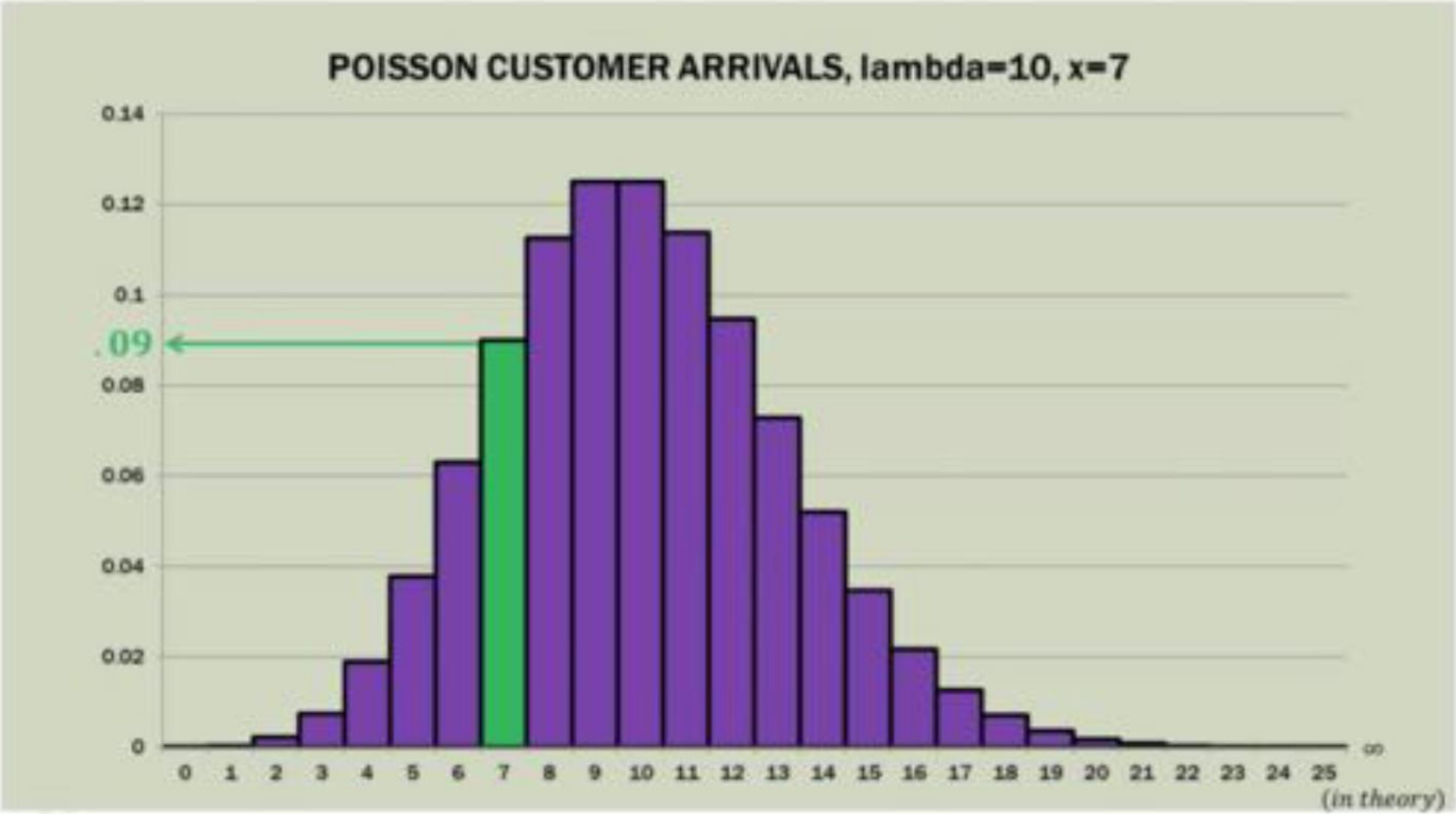
λ = 10

e = 2.718282

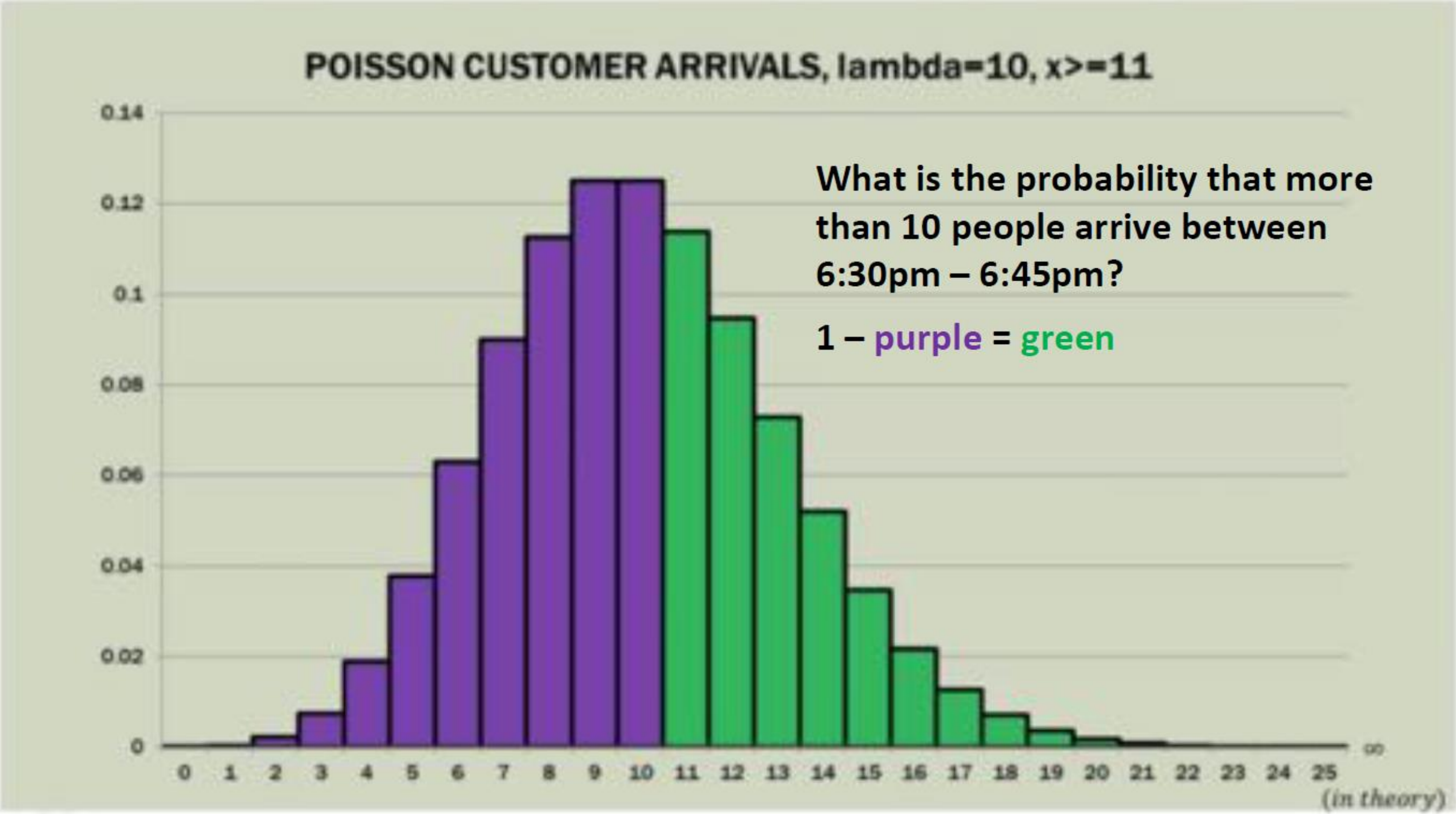
$$P(7) = \frac{10^7 e^{-10}}{7!}$$

$$P(7) = 0.09 \ or \ 9\%$$

Forward School



Forward School



What is the probability that more than 10 people arrive between 6:30pm – 6:45pm?

1 – purple = green

Forward School

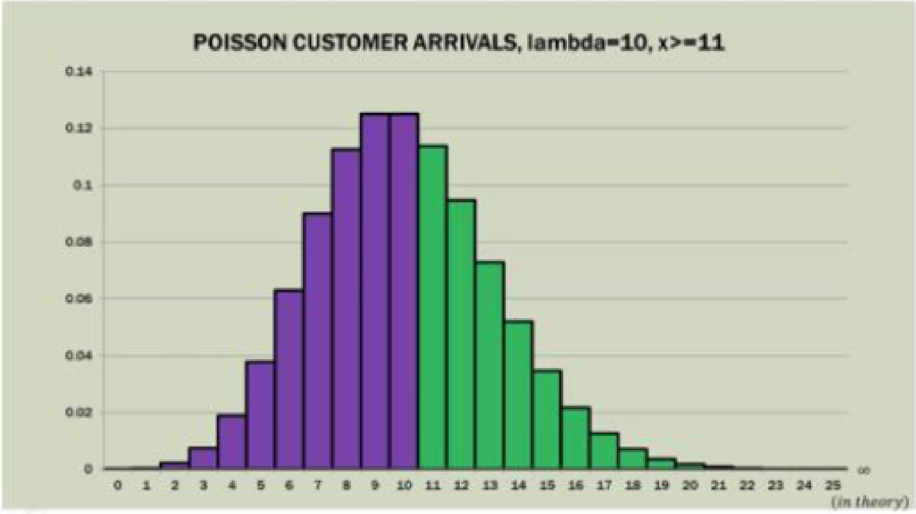
Probability of more than 10 customers

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$1 - P(10) = 1 - \frac{10^{10}e^{-10}}{10!} = 1 - 0.583 = 0.417$$

What is the probability that more than 10 people arrive between 6:30pm – 6:45pm?

$x = 10$
$\lambda = 10$
e = 2.718282



Forward School

Answer Summary

- What is the probability that exactly 7 customers enter your line between 6:30pm – 6:45pm?

Answer: 0.09 or 9%

- What is the probability that more than 10 people arrive between 6:30pm – 6:45pm (which means you will probably be on shift later than 7:00pm)?

Answer: 0.417 or 42%

Forward School

Exercise 1

- Let's assume that the manager decides to be nice and you get to close your cashier register at 6:40pm instead of 6:45pm. How many customers do you expect to arrive between 6:30pm and 6:40pm?

$$\frac{10\ customers}{15\ minutes} = \frac{?}{10\ minutes}$$

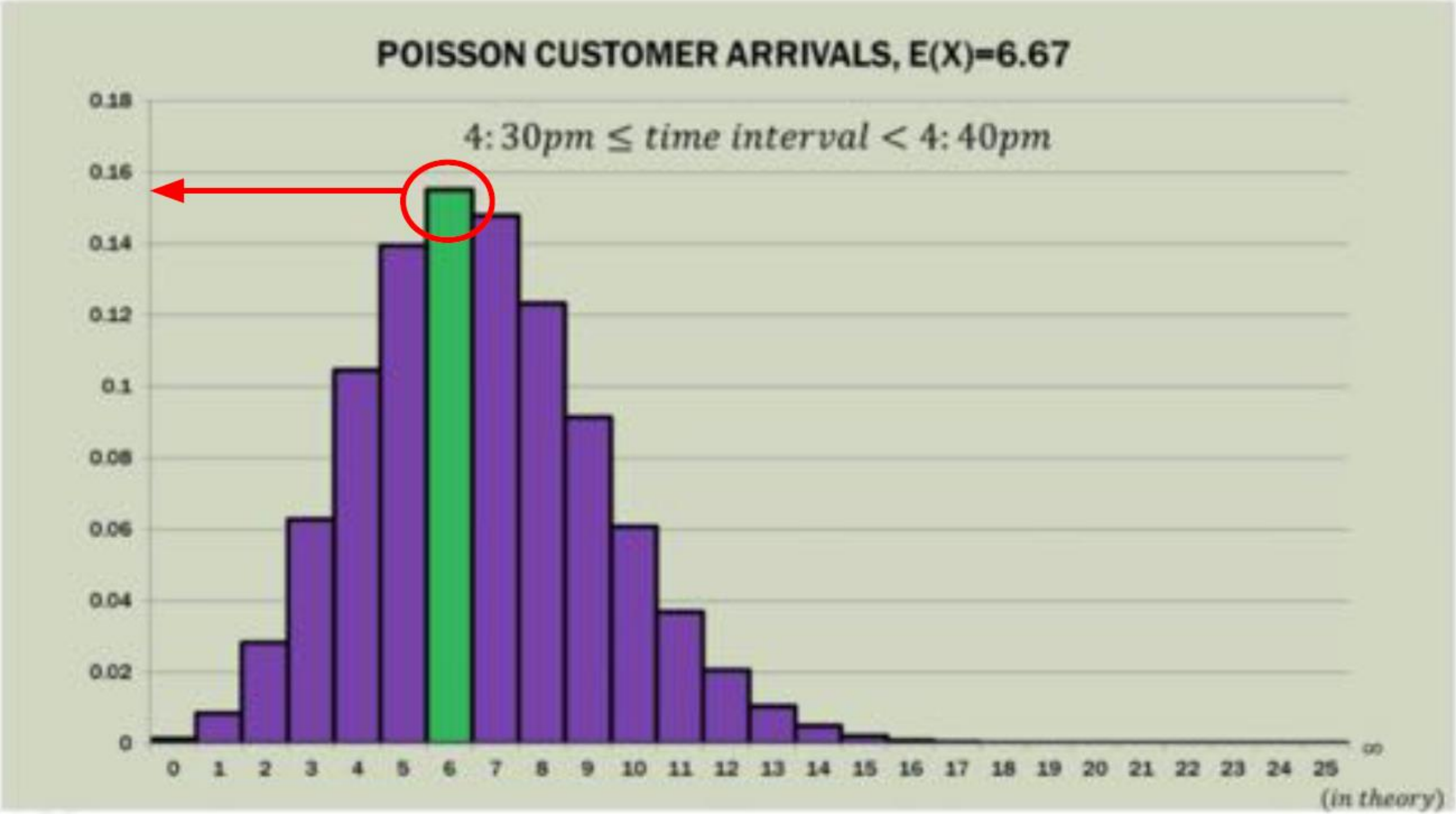
Exercise 1: Solution

- Let's assume that the manager decides to be nice and you get to close your cashier register at 6:40pm instead of 6:45pm. How many customers do you expect to arrive between 6:30pm and 6:40pm?

$$\frac{10\ customers}{15\ minutes} = \frac{?}{10\ minutes}$$

$$\frac{10\ customers}{15\ minutes} = \frac{6.67\ customers}{10\ minutes}$$

Forward School



Forward School

Exercise 2

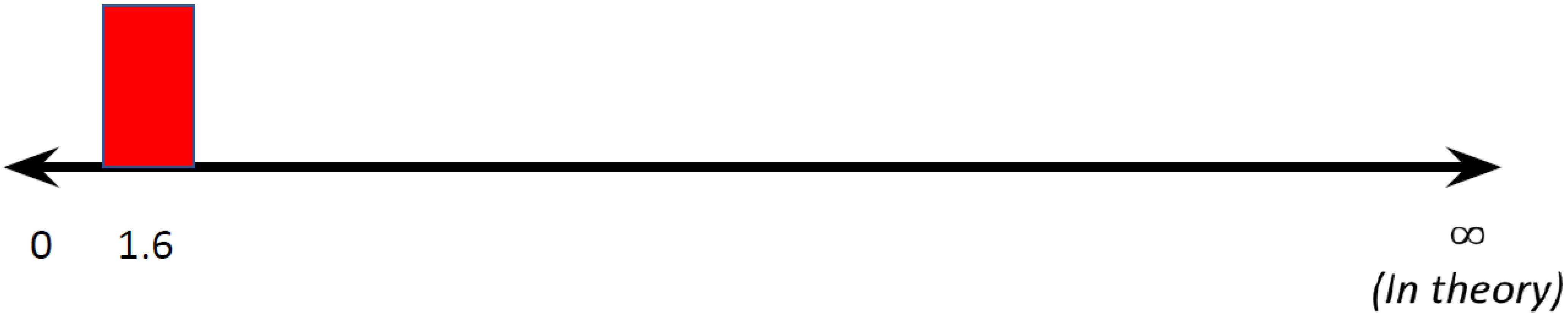
- A bank is interested in studying the number of people who use ATM located outside of its office late at night.
- Based on the data, on average 1.6 customers walk up to the ATM during any 10 minute interval between 9pm and midnight.
- Questions:
  - What is lambda λ for this problem?
  - What is probability of exactly 3 customers using the ATM during any 10 minute interval?
  - What is the probability of 3 or fewer people?

Forward School

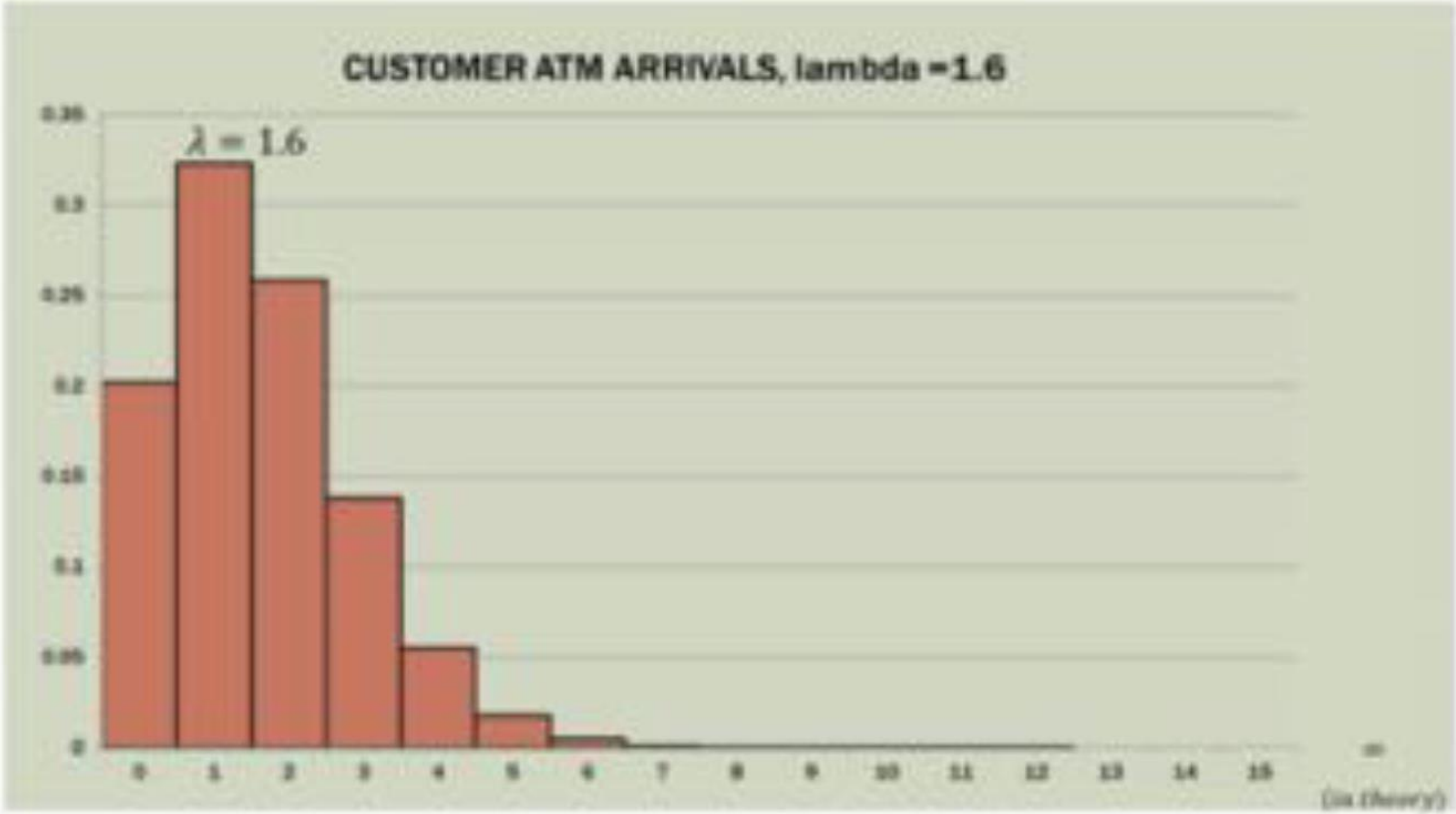
Exercise 2: Solution

Between 9:00 pm – Midnight

Which outcome (#customer) is your expected value/mean?



Forward School


CUSTOMER ATM ARRIVALS, lambda =1.6
λ = 1.6
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
∞
(in theory)

Poisson Distribution

ATM Arrivals

$$\lambda = \frac{\#\ occurrences}{specified\ intervals}$$

$$\lambda = \frac{1.6\ customers}{per\ 10\ minutes} = 1.6$$

What is probability of exactly 3 customers using the ATM during any 10 minute interval?

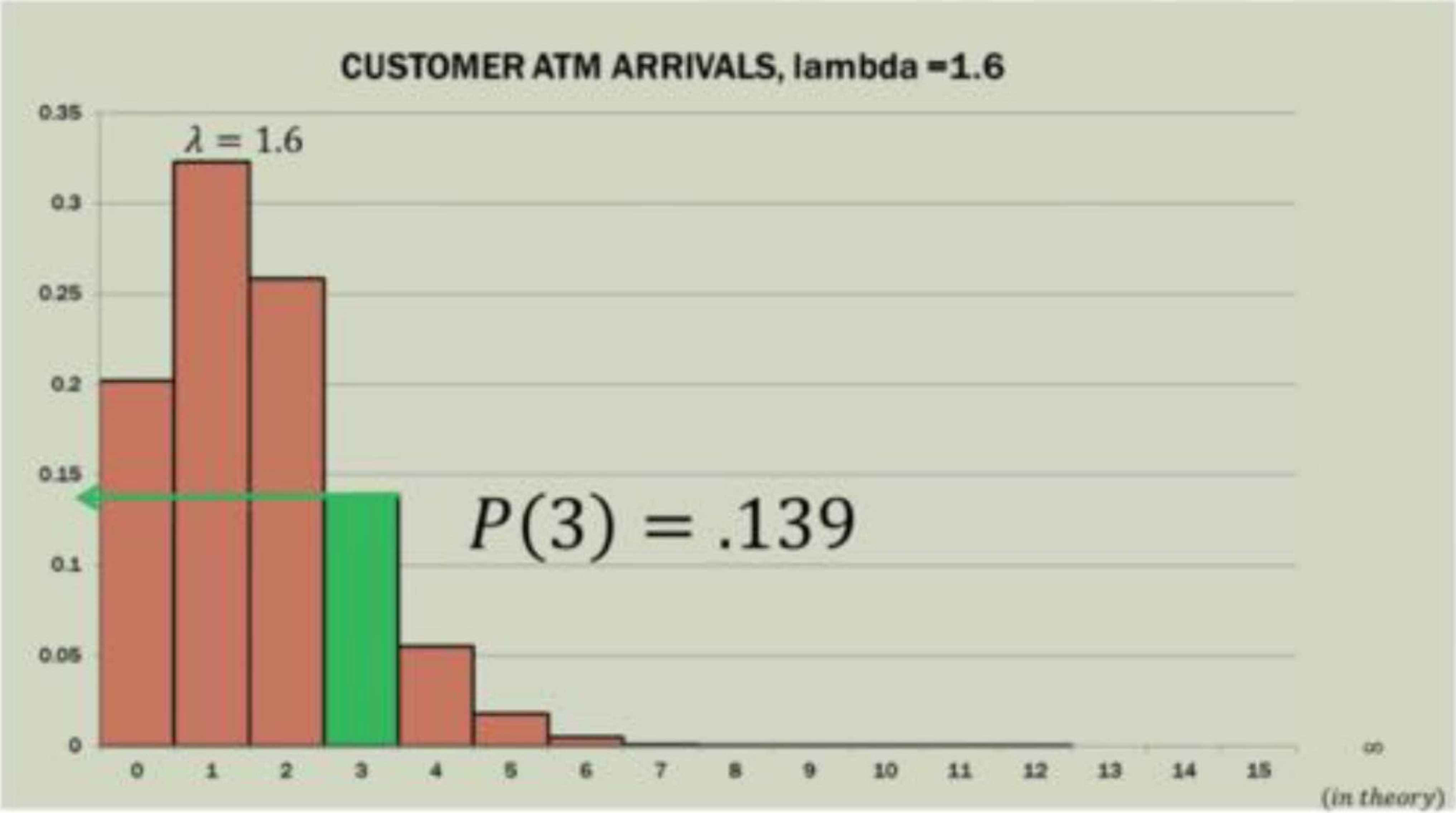
$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} = P(3) = \frac{1.6^3 e^{-1.6}}{3!} = 0.139 \text{ or } 14\%$$

$x = 3$

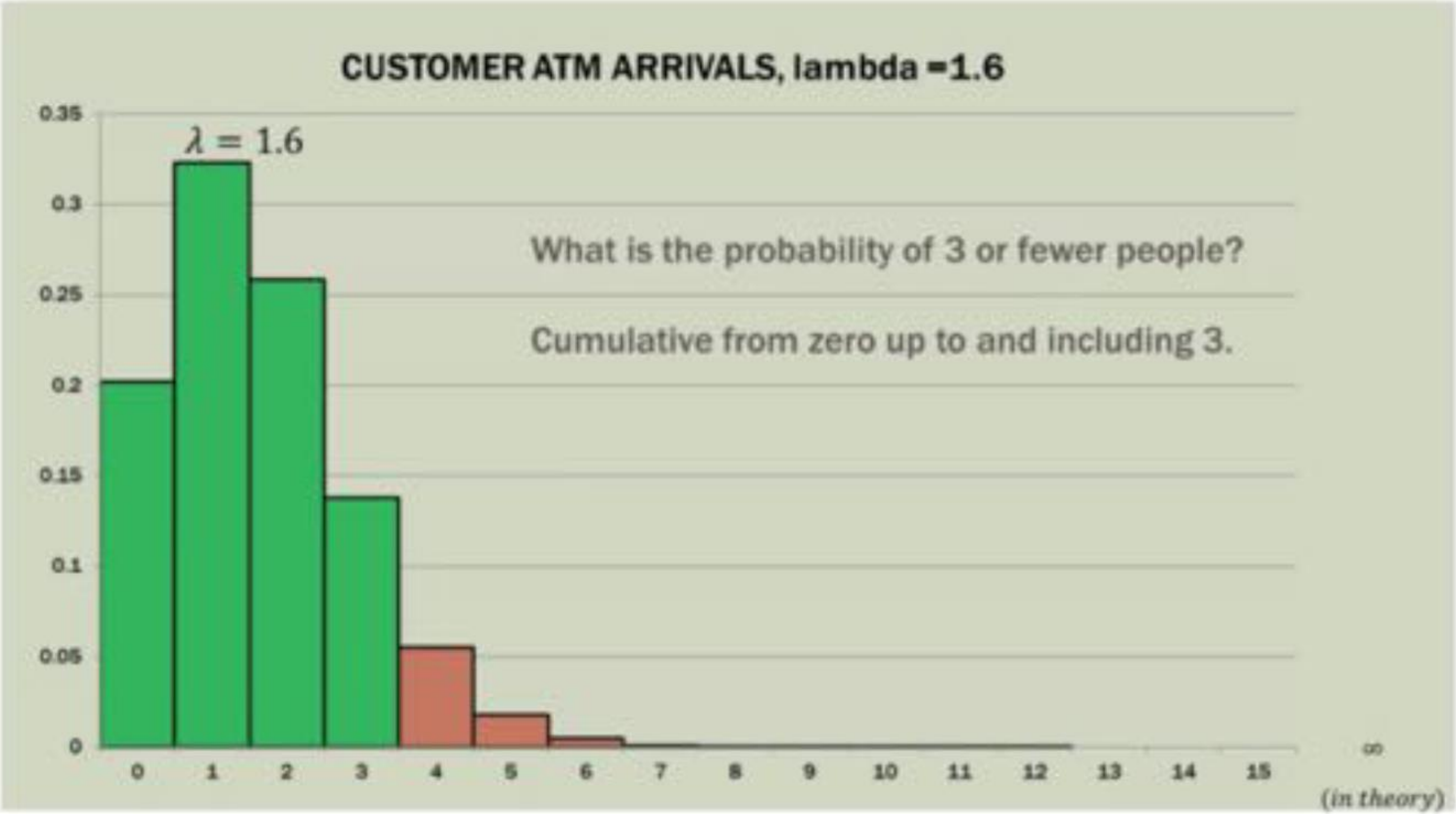
$\lambda = 1.6$

e = 2.718282

Forward School



Forward School



Forward School

Answer Summary

- What is probability of exactly 3 customers using the ATM during any 10 minute interval?

Answer: 0.139 or 14%

- What is the probability of 3 or fewer people?

Answer: 0.9211 or 92%

Forward School

Poisson Distribution in Scipy

- scipy.stats.poisson
- The Poisson distribution is the probability distribution of independent event occurrences in an interval. If λ is the mean occurrence per interval, then the probability of having x occurrences within a given interval is:

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

```
poisson.pmf(k, mu, loc=0)
poisson.cdf(k, mu, loc=0)
poisson.ppf(q, mu, loc=0)
poisson.rvs(mu, loc=0, size=1, random_state=None)
```

Forward School

poisson.pmf(k, mu, loc=0)

- `poisson.pmf` is the Poisson probability mass function or "density"

```
# Calculate p(y) for y=0,1,...,10 when μ=2

from scipy.stats import poisson
Import numpy as np
poisson.pmf(np.arange(0,11),2)

[1] array([1.35335283e-01, 2.70670566e-01, 2.70670566e-01,
1.80447044e-01, 9.02235222e-02, 3.60894089e-02, 1.20298030e-
02, 3.43708656e-03, 8.59271640e-04, 1.90949253e-04,
3.81898506e-05])
```

poisson.cdf(k, mu, loc=0)

- `poisson.cdf` is the cumulative distribution function, P(Y <= y)

```
# Find P(Y <= 6) when μ=2:

> poisson.cdf(6, 2)
[1] 0.9954662
Or
> sum(poisson.pmf(np.arange(0,7),2))
[1] 0.9954662
```

Forward School

Poisson Distribution Table

- # Make a table of the first 10 Poisson probs and cumulative probs when mu=2:

```
import pandas as pd
x1 = poisson.pmf(np.arange(0,10),2)
x2 = poisson.cdf(np.arange(0,10),2)
pd.DataFrame(list(zip(x1,x2)))
```

[118]:

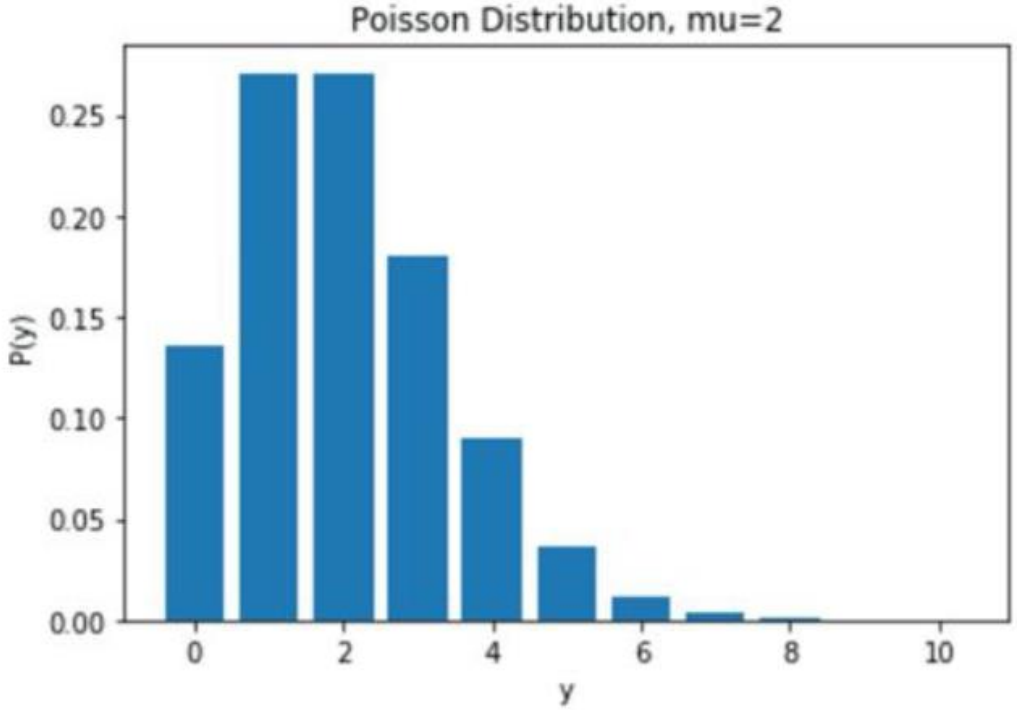
	0	1
0	0.135335	0.135335
1	0.270671	0.406006
2	0.270671	0.676676
3	0.180447	0.857123
4	0.090224	0.947347
5	0.036089	0.983436
6	0.012030	0.995466
7	0.003437	0.998903
8	0.000859	0.999763
9	0.000191	0.999954

Forward School

Plot

- # Plot the probabilities (type="h" makes this particular type of plot # with the vertical lines):

```
import matplotlib.pyplot as plt
%matplotlib inline
x = np.arange(0,11)
y = poisson.pmf(np.arange(0,11),2)
plt.bar(x,y)
plt.title('Poisson Distribution,
mu=2')
plt.ylabel('P(y)')
plt.xlabel('y')
```



Forward School

Exercise 3

- A random variable X has Poisson distribution with mean 7. Find the probability that:
  - X is less than 5
  - X is greater than 10 (strictly)
  - X is between 4 and 16

Forward School