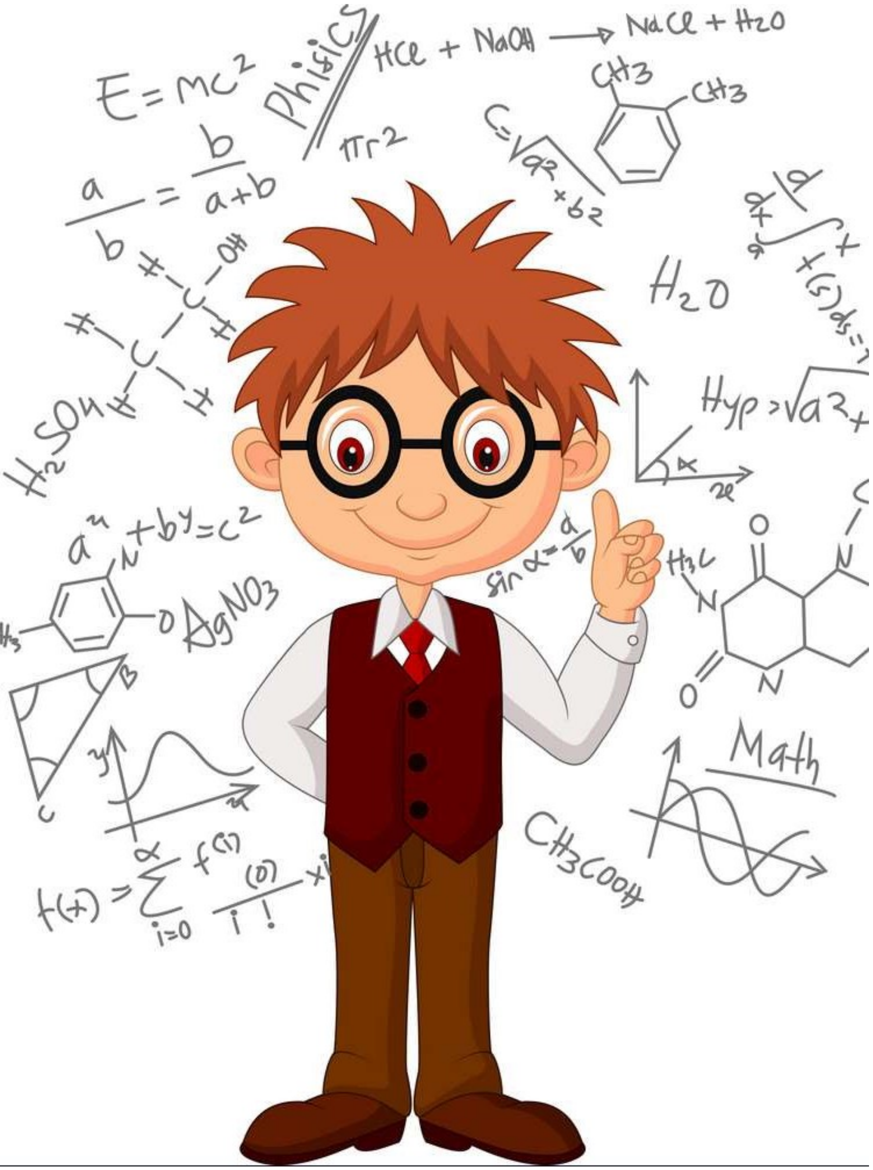
Statistics for Data Analytics

Lecturer: Marina Iantorno
E-mail: miantorno@cct.ie

CCT College Dublin
Computing • IT • Business



In today's class we will cover:

- ❑Poisson distribution

$$P(X=k) = \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k}$$

$p = \frac{\lambda}{n}$

when $n \to \infty$, $p \to 0$.

$$= \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!\,k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n} \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$= \lim_{n\to\infty} \frac{n!}{(n-k)!} \frac{1}{n^k} \boxed{\frac{\lambda^k}{k!} \quad e^{-\lambda}} \quad 1$$

Poisson!

Poisson Distribution

Poisson distribution

The Poisson distribution is another discrete distribution, and it is considered along with the Binomial, one of the most important and frequently used in statistical analysis. The name of this model was given in honoured to his creator, Simeon-Denis Poisson, who was studying the probability of win a gambling game - with very low chances to win – in a large number of tries. He tried to do it using the Binomial distribution (using p as the probability of win and n as the number of trials), but he discovered that there could be another calculation in place that would consider the average numbers of win without taking into consideration the number of trials, because in the end, it was not possible to know how many trials were needed to achieve the goal.

Source of image: https://en.wikipedia.org/wiki/Sim%C3%A9on_Denis_Poisson

Poisson distribution

In simple words, this distribution is used when we consider events in a certain period of time or a physical space where we don't have any defined limit. Some examples when we could use this distribution could be:

- Number of patients who arrive at an emergency room between 6AM and 7AM.
- Number of calls answered by an agent in a call centre.
- Number of holes per meter of cloth.

If you think about it, in all the above cases we have an idea of the maximum number, but we don't know for sure, then, we just consider that there is not

a defined limit. Let's try with an exercise!

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

a) What is the probability of three accidents happening?

b) What is the probability of less than three accidents happening?

c) What is the probability of more than two accidents happening?

d) What is the probability that there are between 2 and 5 accidents inclusive?

e) What is the number of accidents we expect in that corner per day? Calculate Variance and Standard Deviation.

f) If we watch the corner during 5 days, what is the probability of 8 accidents happening?

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

Structure of the random variable

As this is a random variable, we need a definition.

$$X = \underbrace{\text{number of accidents}}_{\textit{Number of elements}} \overbrace{\text{in the corner of CCT}}^{\textit{Specific place}} \underbrace{\text{(per day)}}_{\textit{Time frame/space}}$$

The structure of this variable is ALWAYS

X = number of elements in a specific place within a particular time frame/space

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

Domain of the random variable

As we discussed before, we could also create a table in which we see all the possible values of the variable and their probabilities of occurrence. Again, we need to define the domain. Remember that it includes ALL the possible results when we watch that corner for one day.

Domain: $\{\ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \dots\ \}$

The main characteristic of the Poisson distribution is that there is no limit in the domain. We know that there cannot be a million accidents, but we cannot limit this number to a particular value, therefore we say that there is no limit.

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

Parameters of this distribution

The Poisson distribution has one parameter:

λ = average or expected value

**Note, this number may vary according to the space/time frame.*

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

a) What is the probability of three accidents happening?

X = number of accidents that happens at the CCT corner (in one day)

λ = 2

P(X = 3) =

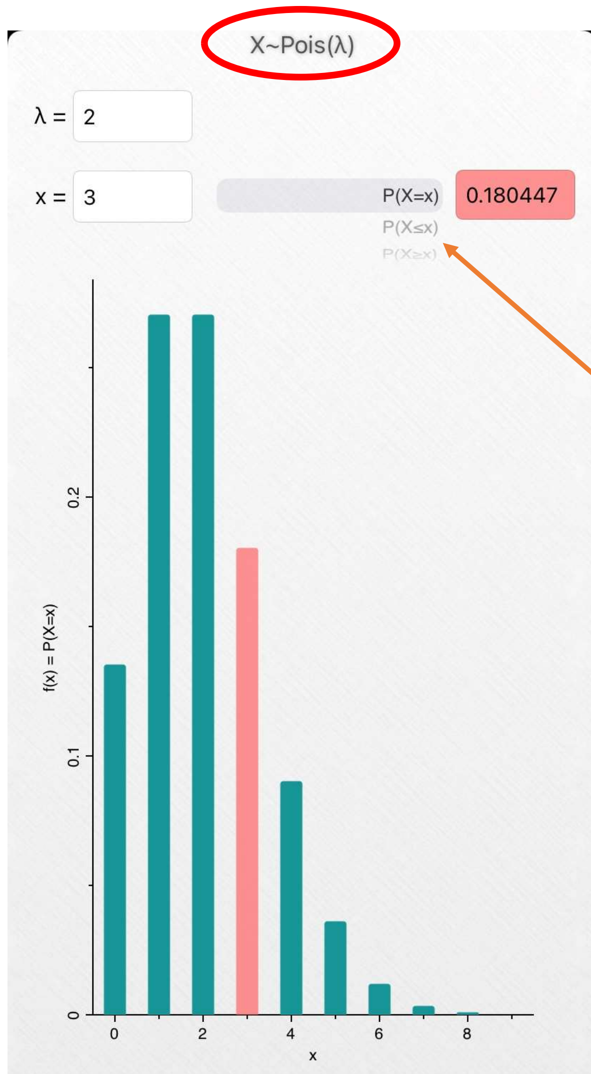
Formula

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

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**...**

We can look for this probability on the Probability Distributions App.

We fill these two gaps



We see the result here

We choose

=

≥

≤

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

What is the probability of three accidents happening?

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

$P(X = 3) = 0.1805$

Formula

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

0	1	2	3	4	5	6	7	...

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

b) What is the probability of less than three accidents happening?

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

$P(X < 3) =$ **Less than 3 doesn't include 3!**

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**...**

We can look for this probability on the App!

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

b) What is the probability of less than three accidents happening?

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

$P(X < 3) = F(2) = \boxed{0.6767}$

Less than 3 doesn't include 3!

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

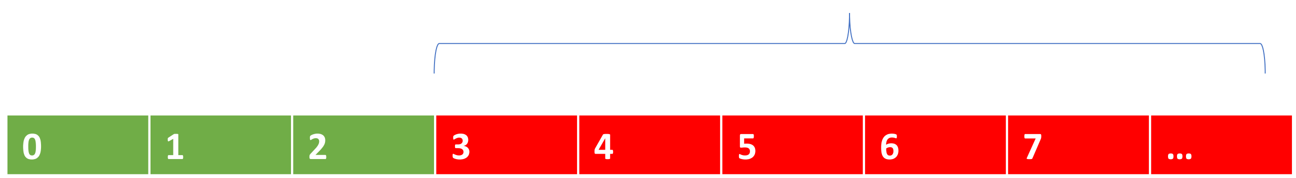
c) What is the probability of more than two accidents happening?

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

$P(X > 2) =$

More than 2 doesn't include 2!



Here we have to be very careful! This is a linguistic problem!

Poisson Distribution

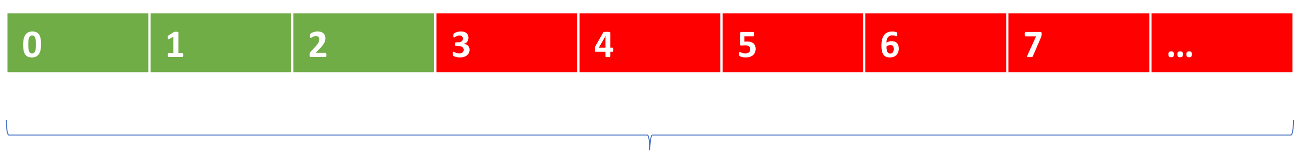
At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

c) What is the probability of more than two accidents happening?

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

$P(X > 2) = 1 - F(2) = 1 - 0.6767 = 0.3233$



If we summed all the probabilities, the result would be 1, but here we don't have a limit, therefore we need to work with the complement probability!

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

d) What is the probability that there are between 2 and 5 accidents inclusive?

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

$P(2 \leq X \leq 5) =$

We have two possible ways to solve it. Any ideas?

0	1	2	3	4	5	6	7	...

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

d) What is the probability that there are between 2 and 5 accidents inclusive?

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

Option 1:

$P(2 \leq X \leq 5) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) = 0.5775$

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**...**

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

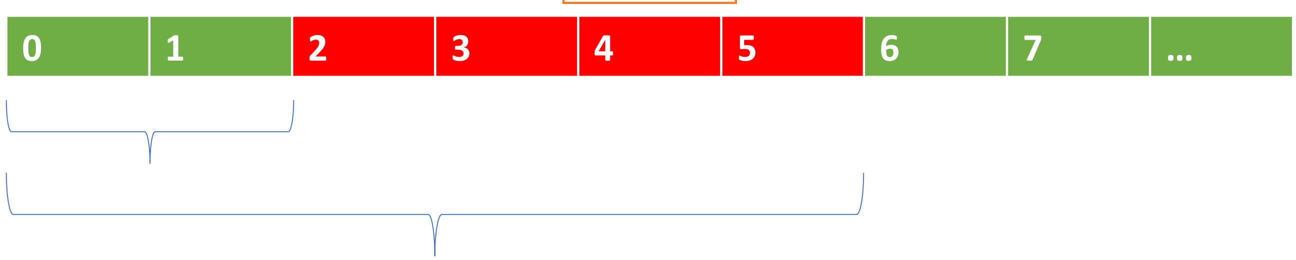
d) What is the probability that there are between 2 and 5 accidents inclusive?

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

Option 2:

$P(2 \leq X \leq 5) = F(5) - F(1) = 0.9835 - 0.4060 = \boxed{0.5775}$



Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

e) What is the number of accidents we expect in that corner per day? Calculate Variance and Standard Deviation.

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

The expected value is the parameter of this distribution, therefore:

$E(X) = \lambda \rightarrow \boxed{E(X) = 2}$

Answer: We expect 2 accidents per day in that corner

Poisson Distribution

e) What is the number of accidents we expect in that corner per day? Calculate Variance and Standard Deviation.

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

The variance of this distribution is also lambda!

$VAR(X) = \lambda \rightarrow \boxed{VAR(X) = 2 \text{ accidents}^2}$

The Standard Deviation is the square root of lambda

$SD(X) = \sqrt{\lambda}$

$\boxed{SD(X) = \sqrt{2 \text{ accidents}^2} = 1.41 \text{ accidents}}$

Answer: We expect 2 accidents per day in that corner + - 1.41 accidents

Poisson Distribution

e) What is the number of accidents we expect in that corner in 3 days?

X = number of accidents that happens at the CCT corner (3 days)

$\lambda = 2$

1 day -----   2 accidents

3 days ----- $\lambda = (3*2) / 1 = 6$ accidents

$E(X) = \lambda \rightarrow E(X) = 6$ accidents

If we changed the time/space, we change lambda.

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day.

f) If we watch the corner during 5 days, what is the probability of 8 accidents to happen?

This sentence will modify all our exercise. Any ideas?

X = number of accidents that happens at the CCT corner (in one day)

$\lambda = 2$

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

f) If we watch the corner **during 5 days**, what is the probability of 8 accidents to happen?

If the time frame or space changes, so does lambda!

X = number of accidents that happen at the CCT corner (in five days)

1 day ———— 2 accidents

5 days ———— $\lambda = (5 * 2)/ 1 = 10$ ← This is our new lambda

P(X = 8) =

Poisson Distribution

At the corner of the CCT College there are, in average, two car accidents a day. If we watch the corner any random day:

f) If we watch the corner **during 5 days**, what is the probability of 8 accidents to happen?

If the time frame or space changes, so does lambda!

X = number of accidents that happens at the CCT corner (in five day)

$\lambda = 10$

$P(X = 8) = \boxed{0.1126}$

THAT'S ALL FOR TODAY

THANK YOU