12th December, 2022

DSML : CC Maths

Probability 9 - Distributions - 4

Recap: (a) Probability theory.
(b) Bayes' theorem.
(c) Combinatorics.
(d) Descriptive statistics.
(e) Binomial Distribution.
(f) Gaussian Distribution
(g) Poisson Distribution.
(h) Geometric Distribution.

Class starts @ 9:05 p.m.

Today: (a) Exponential Distribution.
(b) Log-normal Distribution.
(c) Problem solving.

Before we begin,

(a) If X is a geometric random variable with probability of success p, $E[X] = 1/p$.

* scipy.stats.geom.expect gives wrong answers for $p < 0.03$.

Q] You go to a party with 750 guests. What is the probability that exactly one guest has the same birthday as you?

* $P[\text{same b'day}] = \dfrac{1}{365}$

binom.pmf (k = 1, n = 750, p = 1/365)

Poisson approximation to Binomial.

If for the binomial distribution,

· $n \geq 30$, · $p \leq 0.05$.

then instead of:

binom.pmf(n, k, p) $\longrightarrow$ poisson.pmf(k, mu).

$$E[X] = 1 \;—\; p$$
$$n \;—\; ?$$

$x = n \times p$.

$$\lambda = n \cdot p$$
$$mu = n \cdot p.$$

Suppose we toss a coin once every 10 minutes.
Let $p = 0.2778$ (Prob. of heads).

binom.pmf(k, n, p)
binom.cdf(k, n, p)
binom.expect(args=(n,p))

Q1) What is the probability of getting 1 heads in 30 minutes?

$$P[X=1] = \text{binom.pmf}(k, n, p)$$

0.4346 · k = 1 · n = 3 · p = 0.2778

Q2) Expected no. of heads in 30 mins?

binom.expect(n = 3, p = 0.2778)

0.8334.

Q3] Prob. of getting 1 H in 90 mins? → 0.185

Q4] Expected no. of heads in 90 mins? → 2.5.

Summary

	1 coin, 10 mins ($p = 0.27778$)	1 coin, 1 min ($p = 0.02778$)	10 coins / min ($p = 0.002778$)
30 mins	$P[X=1] = 0.4346$ $E[X] = 0.8334$	$P[X=1] = 0.368$ $E[X] = 0.8334$	$P[X=1] = 0.368$ $E[X] = 0.8334$
90 mins	$P[X=1] = 0.185$ $E[X] = 2.5$	$P[X=1] = 0.203$ $E[X] = 2.5$	$P[X=1] = 0.203$ $E[X] = 2.5$

$\rightarrow \lambda = \underline{0.8334}$

$n = 3$ $n = 30$ $n = 300$

$n = 9$ $n = 90$ $n = 900$

$\lambda = \underline{2.5}$

Suppose we toss a coin once every minute.

$p = 0.02778.$

$P[X=1]$ in 30 mins : 0.3681

$E[X]$ in 30 mins : 0.8339.

$P[X=1]$ in 90 mins : 0.2037

$E[X]$ in 90 mins : 2.5.

Suppose we toss a coin 10 times per minute.

$p = 0.002778.$

$P[X=1]$ in 30 mins : $\underline{0.362}$.

$E[X]$ in 30 mins : 0.8334.

$n = 300$

$k = 1$

$p = 0.002777$

$P[X=1]$ in 90 mins : $\underline{0.205}$

$E[X]$ in 90 mins : 2.5

$n = 900$

$k = 1$

$p = 0.00277$

Poisson : poisson.pmf(k , mu)
poisson.cdf (k , mu)
poisson.expect (args = (mu,))

* Football matches have an average rate of goals as 2.5 in 90 mins.

$\frac{2.5}{3}$

Q) P[x = 1] in 30 mins.

poisson.pmf(k = 1 , mu = 2.5/3)

Q] P [x = 1] in 90 mins.

poisson.pmf (k = 1 , mu = 2.5).

Poisson distribution

`poisson.pmf(k, mu)` $P[X=k] = \frac{\lambda^k e^{-\lambda}}{k!}$

You receive 240 messages per hour on average - assume Poisson distributed.

Q1) What is the average or expected number of messages in 30 seconds?

1 hour (3600 seconds) — 240 messages

30 seconds — ?

$$\frac{30 * 240}{3600} = 2$$

Q2) What is the probability of one message arriving over a 30 second time interval?

If we consider 30 seconds as one unit interval, then $\lambda = 2$

$P[X=1] =$ `poisson.pmf(k=1, mu=2)` $= 0.27$

$$P[X=1] = \frac{(2)^1 e^{(-2)}}{1!} = 0.27$$

Q3) What is the probability that there are no messages in 15 seconds? $\lambda = \frac{15 * 240}{3600} = 1$

$P[X=0] =$ `poisson.pmf(k=0, mu=1)` $= 0.367$

$$P[X=0] = \frac{(1)^1 e^{(-1)}}{1!} = 0.367$$

Q4) What is the probability that there are 3 messages in 20 seconds? $\lambda = \frac{20 * 240}{3600} = 1.33$

$P[X=3] =$ `poisson.pmf(k=3, mu=1.33)` $= 0.104$

$$P[X=3] = \frac{(1.33)^3 e^{(-1.33)}}{3!} = 0.104$$

Poisson distribution

poisson.pmf(k, mu)

$$P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

You receive 240 messages per hour on average - assume Poisson distributed.

Q1] Average wait time between two messages?

15 s

Q2] Average # of messages / sec? → $\frac{1}{15} = 0.067 s$

Q3] What is the prob. of having 0 messages in 10 seconds? $mu = \frac{10}{15}$.

→ 0.5134.

$$P[X=0] = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda}$$

Q4] What is the probability of waiting for more than 10s for the next message?

Let T denote the time to wait for the next message.

$mu = \frac{10}{15}$. $P[X=0]$

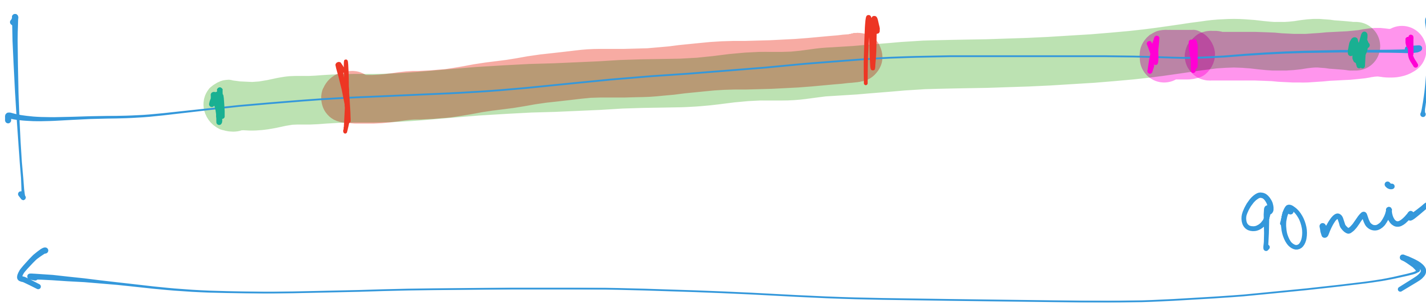
Poisson distribution

`poisson.pmf(k, mu)` $P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$

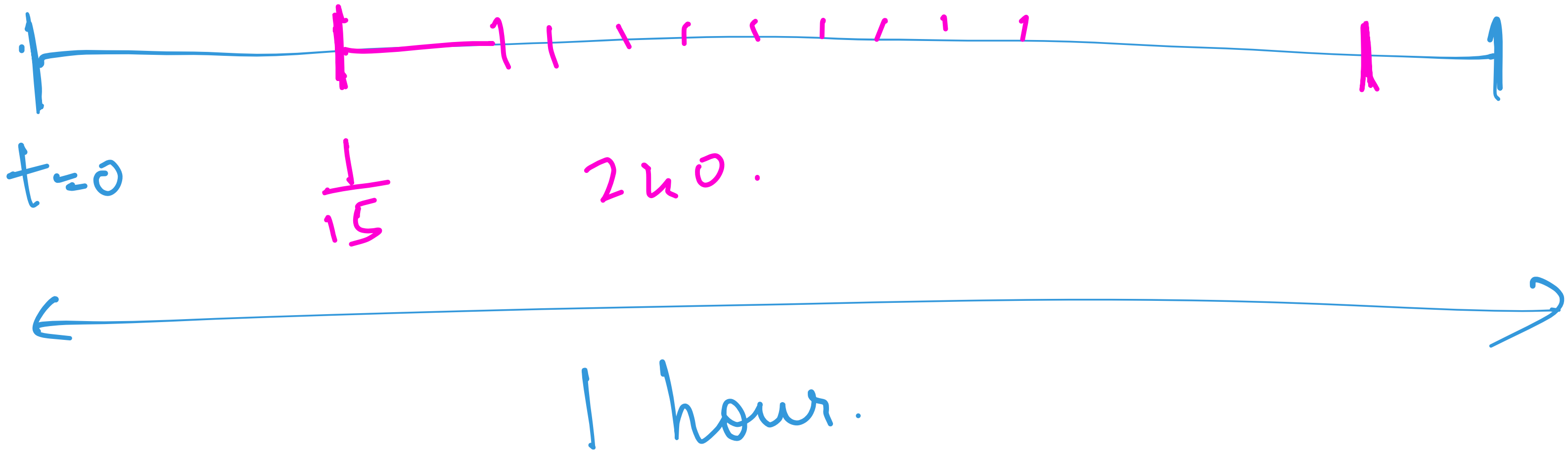
You receive 240 messages per hour on average - assume Poisson distributed.

Possion R.V. X → Counts # of occurrences in a certain time period.

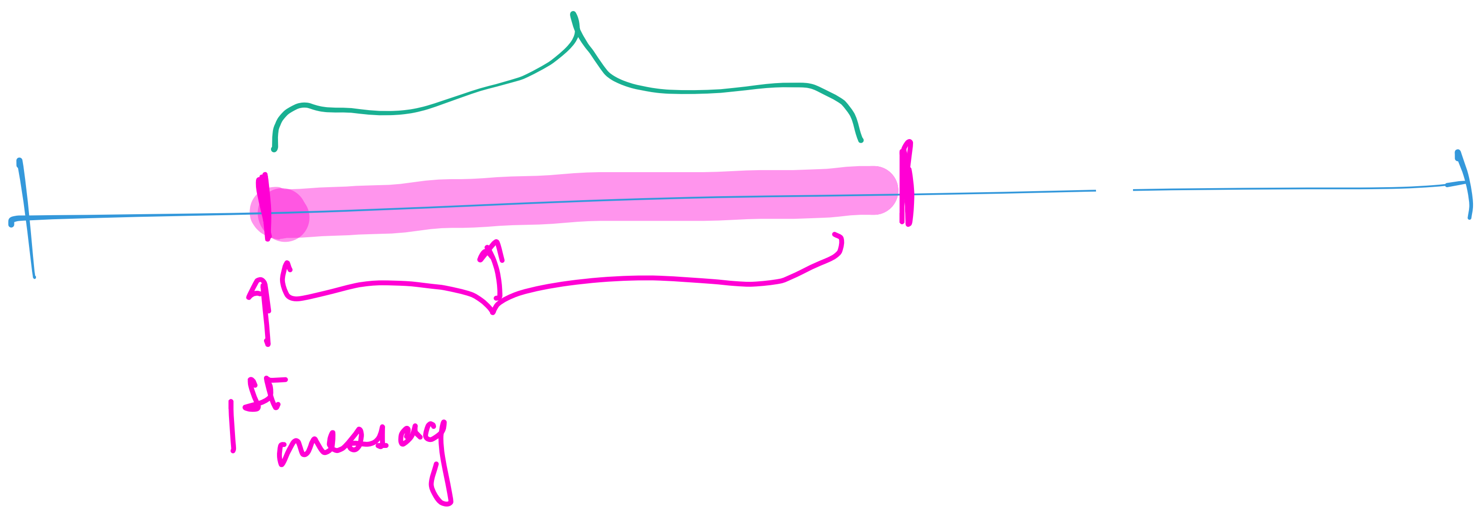
Exponential R.V. T → measures the time interval between two occurrences.



90 mins football match.



t=0
1/15
240.
1 hour.



1st message

Poisson distribution

`poisson.pmf(k, mu)` $P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}$

You receive 240 messages per hour on average - assume Poisson distributed.

Q] What is the probability of waiting for less than or equal to 10 seconds?

$P[T > 10] = 0.5134.$

$P[T \leq 10] = 1 - 0.5134 = 0.4865.$

You are working as a data engineer who has to resolve any bugs/ failures of machine learning models in production

The time taken to debug is exponentially distributed with mean of 5 minutes

$P[T \leq x] = 1 - e^{-x\lambda}$

```
expon.cdf(x, scale)
```

Scale = average wait time between two events

Q1] Prob. of debugging in 4–5 min.

$$P[4 \leq T \leq 5]$$

$$= \text{expon.cdf}(x = 5, \text{scale} = 5) - \text{expon.cdf}(x = 4, \text{scale} = 5)$$

Q2] Prob. of needing more than 6 minutes to debug:

$$P[T > 6] = 1 - P[T \leq 6]$$

$$= 0.3012.$$

You are working as a data engineer who has to resolve any bugs/ failures of machine learning models in production

The time taken to debug is exponentially distributed with mean of 5 minutes

$P[T \le x] = 1 - e^{-x\lambda}$

```
expon.cdf(x, scale)
```

Q] Given that you have already spent 3 minutes, what is the prob. of needing more $\underbrace{\text{than 9 minutes}}_{\text{"six more minutes"}}$?

$$P[T > 9 \mid T > 3] = \frac{P[T > 3 \text{ and } T > 9]}{P[T > 3]}$$

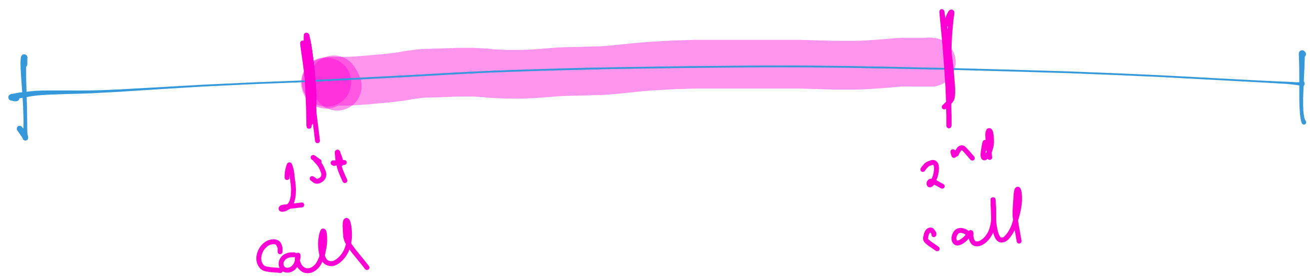
$$\frac{P[T > 9]}{P[T > 3]} = \frac{1 - P[T \le 9]}{1 - P[T \le 3]} = 0.3011$$

Memoryless property:

The fact that you took 3 mins. so far does not affect how much more time you need to debug!!

A call center gets 3.5 calls per hour (average).

Q] Calculate the probability that the next call will arrive at least 30 minutes after the previous call.



T

$P[T \geq 30].$

$1 - \text{expon.cdf}(x = 30, \text{scale} = 1/0.0583)$

$\approx 0.1734.$

1 hour (60 mins) 3.5

1 minute ?

$\frac{3.5}{60} \approx \underline{0.0583}.$

3.5 calls per hour.

$$\frac{3.5}{1} \quad - \quad \frac{60}{?} \text{ minutes.}$$

scale - avg. time between 2 calls.

$$3.5x = 60$$

$$x = \frac{60}{3.5} = 17.14.$$

A call center gets 3.5 calls per hour (average)

Q] Calculate the probability that the next call will arrive at least 30 minutes after the previous call.

1 interval – 30 mins.

1 hour – 2 30 minute intervals.

$2 - 3.5 = \frac{3.5}{2}$.

$1 - ?\ x$

$P[T > 1] = 1 - P[T \leq 1]$

Binomial distribution:

$$p(x=k) = {}^nC_k \cdot p^k (1-p)^{n-k}.$$

$$E[x] = \sum_{k=0}^{n} \underline{k} \cdot P[x=k]$$

Probability of $\underline{k}$ successes in n trials.

$$E[x] = \underbrace{\sum_{k=0}^{n} k \cdot {}^nC_k \, p^k (1-p)^{n-k}}.$$

$$\downarrow$$

$$n \cdot p.$$

Geometric Distribution.

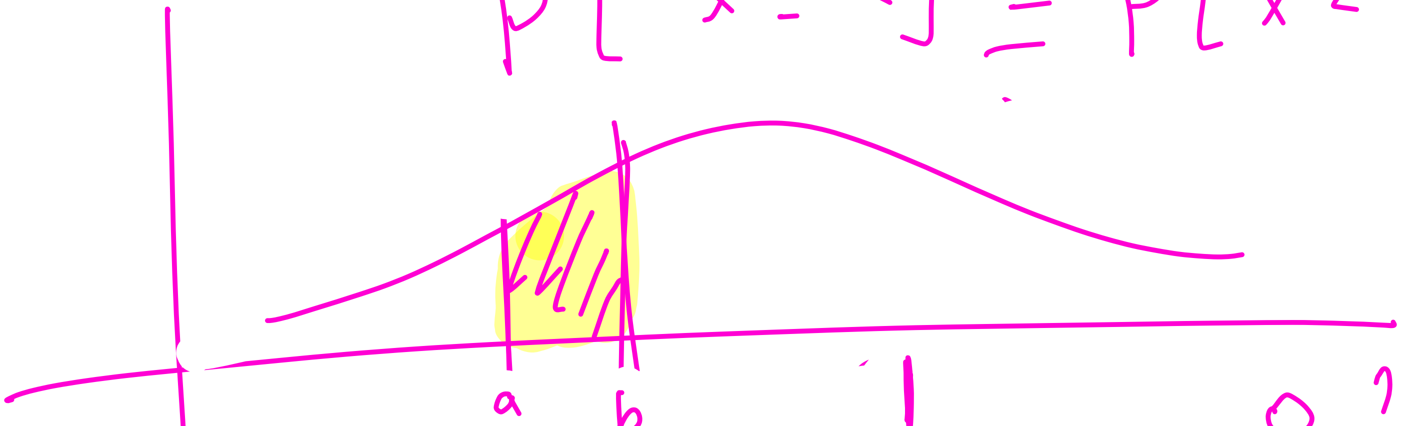
$$P[X = k] = (1-p)^{k-1} p.$$

$$E[X] = \sum_{k=1}^{\infty} k \cdot (1-p)^{k-1} \cdot p.$$

$$\downarrow$$

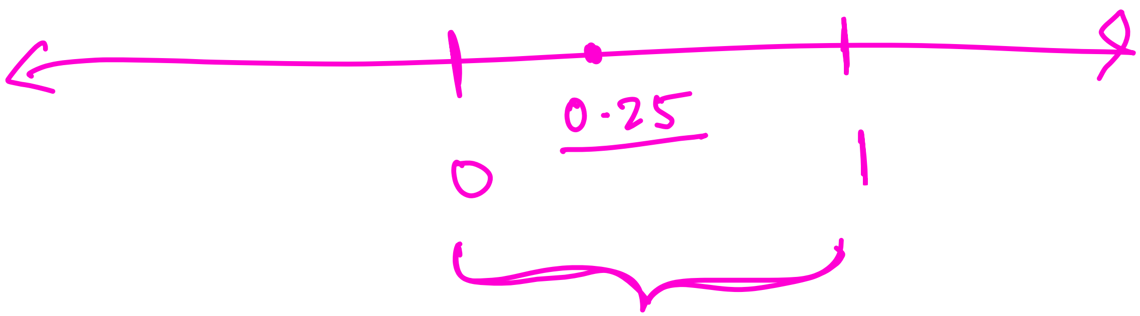
$$\frac{1}{p}.$$

$P[X \leq a] = P[X < a]$.



$\dfrac{1}{\infty} = \underline{\underline{0}}$?

$cdf(b) - cdf(a)$



$$\mathrm{CDF}(a) = \int_{-\infty}^{a} \mathrm{pdf}(x)\,dx$$

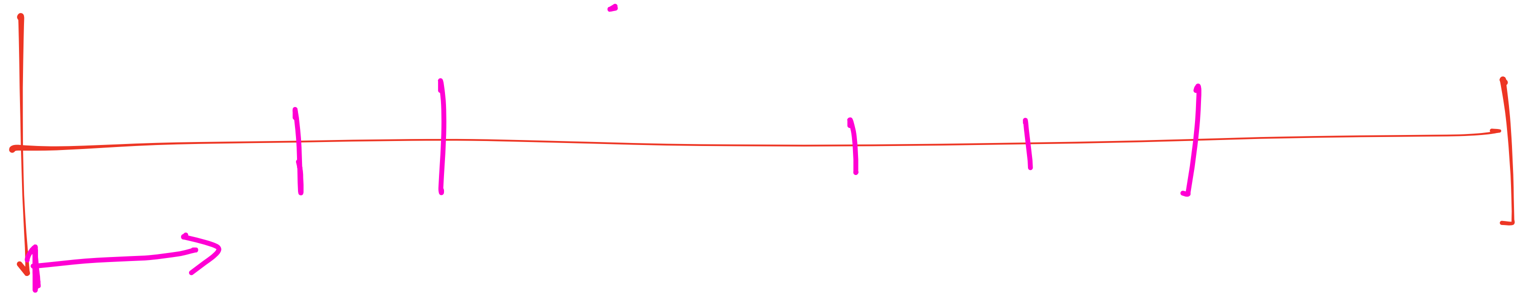
$$\mathrm{CDF}(B) - \mathrm{CDF}(A)$$

$$\int_{-\infty}^{b} \underbrace{\mathrm{pdf}}\,dx - \int_{-\infty}^{a} \underbrace{\mathrm{PDF}}\,dx.$$

$$= \int_{a}^{b} \mathrm{pdf}(x)\,dx.$$

$\frac{240}{6}$ messages — 60

? — 10



0

1] Poisson R.V.

* Probability of getting 0 messages in 10 minutes.

2T , 1H $P(H) = p$

$$P[X=2] = {}^{3}C_2 \cdot p^2 \cdot (1-p)^1$$