Poisson distribution

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

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) $\qquad$ 240 messages $\qquad \dfrac{30 * 240}{3600} = 2$

30 seconds $\qquad$?

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)` $= 0.27$

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

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

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

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

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

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

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

Poisson distribution

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

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

Q1) What is the average time to wait between two messages? $\dfrac{3600}{240} = 15$ seconds

1 hour (3600 seconds) ✕ 240 messages

? 1

Q2) What is the average number of messages per second? $\lambda_1 = \dfrac{1}{15} = 0.067$ per second

Q3) What is the probability of having no messages in 10 seconds? $\lambda_{10} = \dfrac{10}{15} = 0.67$ per 10-seconds

$$P[X = 0] = \dfrac{\lambda_{10}^0 e^{-\lambda_{10}}}{0!} = e^{-\lambda_{10}} = e^{-10\lambda_1} = 0.5134$$ $\lambda_{10} = 10 * \lambda_1$

$P[X = 0] =$ `poisson.pmf(k=0, mu=10/15)` $\lambda_t = t * \lambda_1$

Q4) What is the probability of waiting for more than 10 seconds for the next message?

Let T denote the time to wait for the next message

$$P[T > 10] = e^{-10\lambda_1} = 0.5134$$

Q5) What is the probability of waiting less than or equal to 10 seconds?

$$P[T \leq 10] = 1 - e^{-10\lambda_1} = 0.4865$$

```
from scipy.stats import expon
```
$P[T \leq x] = 1 - e^{-x\lambda}$

$P[T \leq 10] =$ `expon.cdf(x=10, scale=15)` $= 0.4865$

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

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

`expon.cdf(x, scale)`

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

Q1) Find the probability of debugging in 4 to 5 minutes

$P[4 < T < 5] =$ `expon.cdf(`*x=5, scale=5*`)` − `expon.cdf(`*x=4, scale=5*`)` $= 0.0814$

Q2) Find the probability of needing more than 6 minutes to debug

$P[T > 6] =$ `1` − `expon.cdf(`*x=6, scale=5*`)` $= 0.3012$

Q3) Given that you have already spent 3 minutes, what is the probability of needing more than 9 minutes

$$P[T > 9 \mid T > 3] = \frac{P[(T > 9) \cap (T > 3)]}{P[T > 3]} = \frac{P[T > 9]}{P[T > 3]} = \frac{1 - \texttt{expon.cdf}(x=9, scale=5)}{1 - \texttt{expon.cdf}(x=3, scale=5)} = 0.3012$$

$$\frac{P[T > 9]}{P[T > 3]} = \frac{e^{-9\lambda}}{e^{-3\lambda}} = e^{-6\lambda} = P[T > 6]$$

Memoryless:

The fact that you took three minutes so far does not affect how much more you might take to debug

Let T_1 denote the amount of time to wait for first goal

Is T_1 discrete or continuous?

T_1 Time taken for first goal from the start of the match

T_2 Time taken for second goal from point the first goal is scored

Under the assumptions of the poisson process, T_1 and T_2 are independent

These are called "Inter arrival times", and are exponentially distributed

Poisson distribution

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

Football matches have an average of 2.5 goals in 90 mins

Q1) What is the probability of having one goal in 30 mins?

90 mins   ✕   2.5

30 mins   ?

$$\lambda = \frac{30*2.5}{90} = 0.833$$

$$P[X=1] = \texttt{poisson.pmf(k=1, mu=0.833)} = 0.362$$

Q2) What is the probability of having one goal in 90 mins? $\lambda = 2.5$

$$P[X=1] = \texttt{poisson.pmf(k=1, mu=2.5)} = 0.205$$

Let us flip a coin like crazy!!!

```
binom.pmf(k, n, p)

binom.expect(args=(n, p))

binom.cdf(k, n, p)
```

Suppose we toss a coin once every 10 minutes.   **The probability of heads is 0.2778**

What is the probability of getting one heads in 30 minutes? $n = 3$ $p = 0.2778$

$P[X = 1] =$ `binom.pmf(k=1, n=3, p=0.2778)` $= 0.4346$

What is the expected number of heads in 30 minutes?

$E[X] = n*p = 3*0.2778 = 0.8334$ $E[X] =$ `binom.expect(args=(3, 0.2778))` $= 0.8334$

What is the probability of getting one heads in 90 minutes? $n = 9$ $p = 0.2778$

$P[X = 1] =$ `binom.pmf(k=1, n=9, p=0.2778)` $= 0.185$

What is the expected number of heads in 90 minutes?

$E[X] = n*p = 9*0.2778 = 2.5$ $E[X] =$ `binom.expect(args=(9, 0.2778))` $= 2.5$

```
binom.pmf(k, n, p)

binom.expect(args=(n, p))

binom.cdf(k, n, p)
```

Once in 10 mins

$p = 0.2778$

30 minutes

$P[X = 1] = 0.4346$

$E[X] = 0.8334$

90 minutes

$P[X = 1] = 0.185$

$E[X] = 2.5$

```
binom.pmf(k, n, p)

binom.expect(args=(n, p))

binom.cdf(k, n, p)
```

Suppose we toss a coin once every minute. **The probability of heads is 0.02778**

What is the probability of getting one heads in 30 minutes? $n = 30$ $p = 0.02778$

$P[X = 1] = $ `binom.pmf(k=1, n=30, p=0.02778)` $= 0.368$

What is the expected number of heads in 30 minutes?

$E[X] = n*p = 30*0.02778 = 0.8334$ $E[X] = $ `binom.expect(args=(30, 0.02778))` $= 0.8334$

What is the probability of getting one heads in 90 minutes? $n = 90$ $p = 0.02778$

$P[X = 1] = $ `binom.pmf(k=1, n=90, p=0.02778)` $= 0.203$

What is the expected number of heads in 90 minutes?

$E[X] = n*p = 90*0.02778 = 2.5$ $E[X] = $ `binom.expect(args=(90, 0.02778))` $= 2.5$

```
binom.pmf(k, n, p)

binom.expect(args=(n, p))

binom.cdf(k, n, p)
```

	Once in 10 mins $p = 0.2778$	Once per min $p = 0.02778$
30 minutes	$P[X = 1] = 0.4346$ $E[X] = 0.8334$	$P[X = 1] = 0.368$ $E[X] = 0.8334$
90 minutes	$P[X = 1] = 0.185$ $E[X] = 2.5$	$P[X = 1] = 0.203$ $E[X] = 2.5$

```
binom.pmf(k, n, p)

binom.expect(args=(n, p))

binom.cdf(k, n, p)
```

Suppose we toss a coin 10 times every minute.   The probability of heads is 0.002778

What is the probability of getting one heads in 30 minutes? $\quad n = 300 \quad p = 0.002778$

$P[X = 1] =$ `binom.pmf(k=1, n=300, p=0.002778)` $= 0.362$

What is the expected number of heads in 30 minutes?

$E[X] = n * p = 300 * 0.002778 = 0.8334 \quad E[X] =$ `binom.expect(args=(300, 0.002778))` $= 0.8334$

What is the probability of getting one heads in 90 minutes? $\quad n = 900 \quad p = 0.002778$

$P[X = 1] =$ `binom.pmf(k=1, n=900, p=0.002778)` $= 0.205$

What is the expected number of heads in 90 minutes?

$E[X] = n * p = 900 * 0.002778 = 2.5 \quad E[X] =$ `binom.expect(args=(90, 0.02778))` $= 2.5$

```
binom.pmf(k, n, p)

binom.expect(args=(n, p))

binom.cdf(k, n, p)
```

	Once in 10 mins $p = 0.2778$	Once per min $p = 0.02778$	Ten times per min $p = 0.002778$
30 minutes	$P[X = 1] = 0.4346$ $E[X] = 0.8334$	$P[X = 1] = 0.368$ $E[X] = 0.8334$	$P[X = 1] = 0.362$ $E[X] = 0.8334$
90 minutes	$P[X = 1] = 0.185$ $E[X] = 2.5$	$P[X = 1] = 0.203$ $E[X] = 2.5$	$P[X = 1] = 0.205$ $E[X] = 2.5$

```
binom.pmf(k, n, p)

binom.expect(args=(n, p))

binom.cdf(k, n, p)
```

Suppose the rate at which "heads" comes is 2.5 in 90 mins

What is the probability of getting one heads in 30 minutes? $\lambda = \dfrac{2.5}{3} = 0.8334$

$P[X = 1] =$ `poisson.pmf(k=1, mu=0.8334)` $= 0.362$

What is the probability of getting one heads in 90 minutes? $\lambda = 2.5$

$P[X = 1] =$ `poisson.pmf(k=1, mu=2.5)` $= 0.205$

What is the expected number of heads in 90 minutes?

$E[X] = n * p = 900 * 0.002778 = 2.5$ $E[X] =$ `binom.expect(args=(90, 0.02778))` $= 2.5$

```
binom.pmf(k, n, p)

binom.expect(args=(n, p))

binom.cdf(k, n, p)
```

	Once in 10 mins $p = 0.2778$	Once per min $p = 0.02778$	Ten times per min $p = 0.002778$	Poisson $\lambda = 2.5$ **per 90 min**
30 minutes	$P[X=1] = 0.4346$	$P[X=1] = 0.368$	$P[X=1] = 0.362$	$P[X=1] = 0.362$
	$E[X] = 0.8334$	$E[X] = 0.8334$	$E[X] = 0.8334$	$E[X] = 0.8334$
90 minutes	$P[X=1] = 0.185$	$P[X=1] = 0.203$	$P[X=1] = 0.205$	$P[X=1] = 0.205$
	$E[X] = 2.5$	$E[X] = 2.5$	$E[X] = 2.5$	$E[X] = 2.5$

A call centre gets 3.5 calls per hour

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

`expon.cdf(x, scale)`

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

Approach 1: Interval = 1 minute

1 hour (60 min)   3.5

1 minute   ?

$$\frac{3.5}{60} = 0.0583$$

$P[T > 30] =$ `1 - expon.cdf(x=30, scale=1/0.0583)` $= 0.1739$

Approach 2: Interval = 30 minute

1 hour (60 min)   3.5

30 minute   ?

$$\frac{3.5}{2} = 1.75$$

$P[T > 1] =$ `1 - expon.cdf(x=1, scale=1/1.75)` $= 0.1739$

Log-Normal distribution

- Consider an example of number of days of hospitalisation

- Most people may be hospitalised for 3 to 5 days

- Quite often, just 1 or 2 days

- But, there can be a few extreme cases in 20 to 30 days

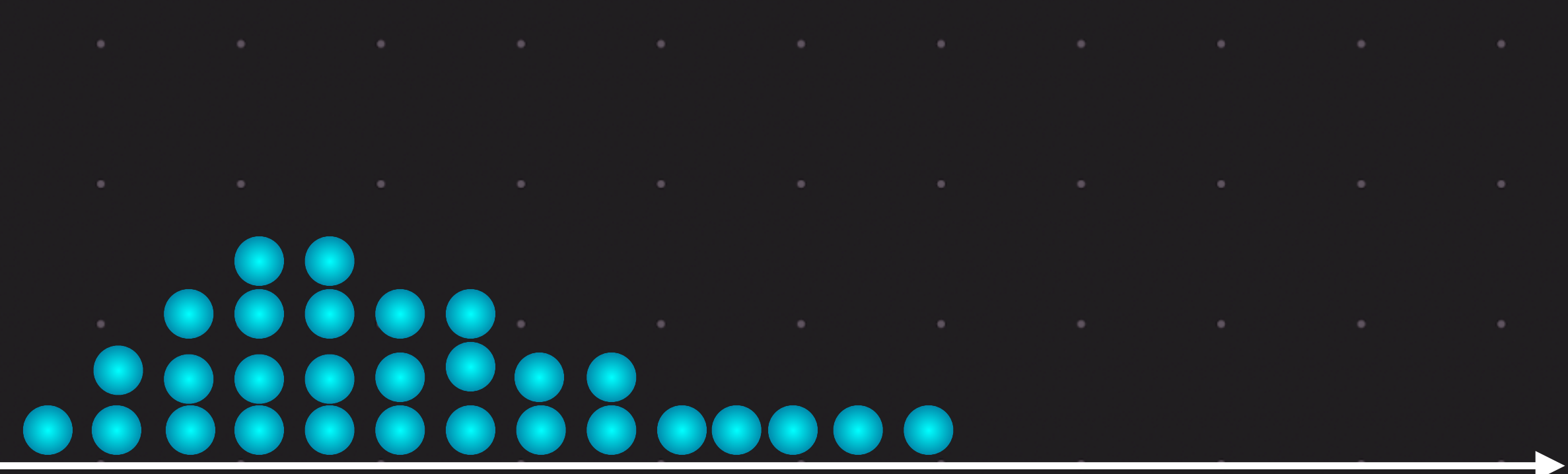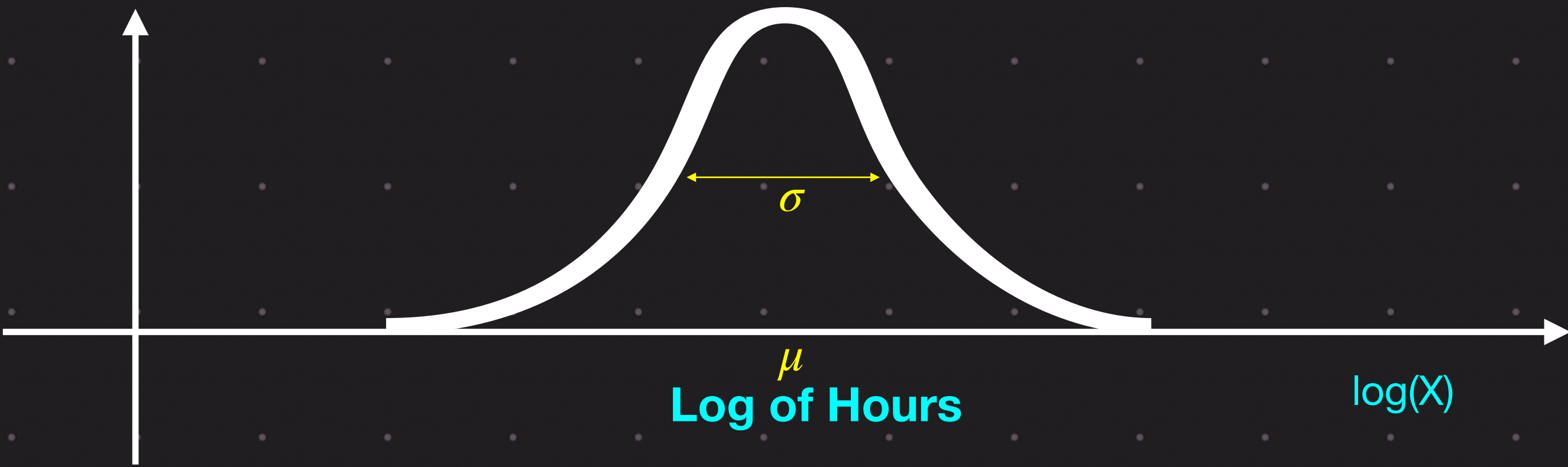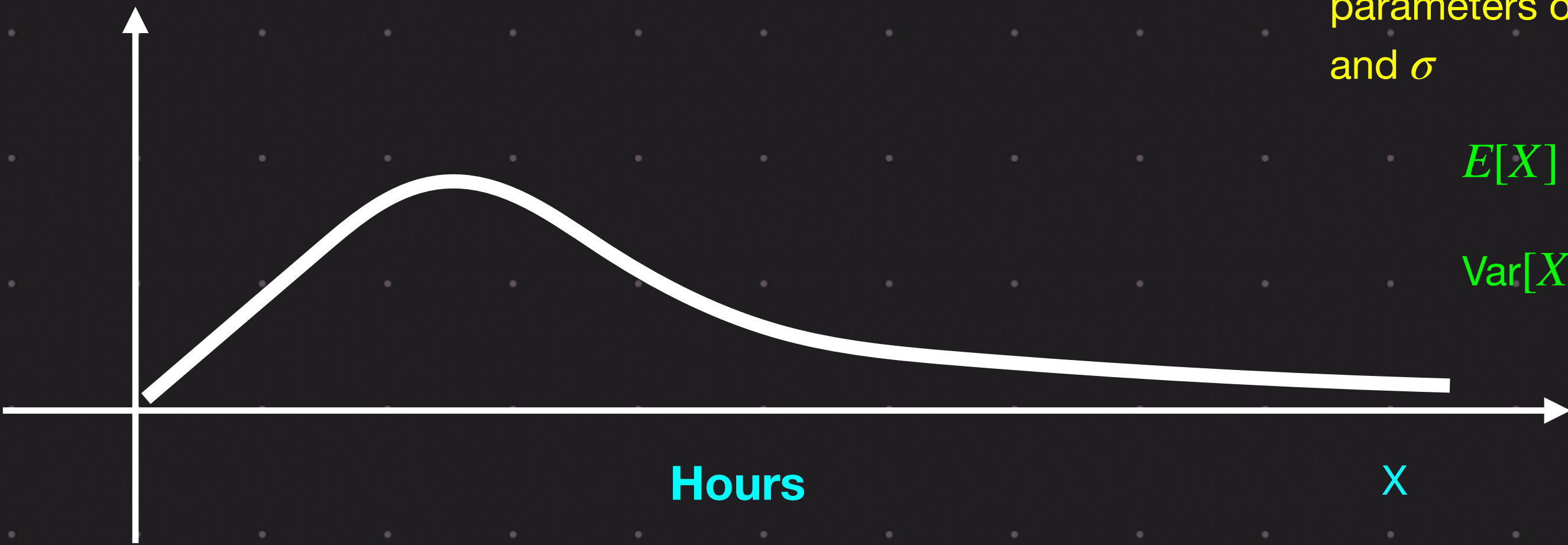- Extremely rare cases of 300 days (coma etc.)
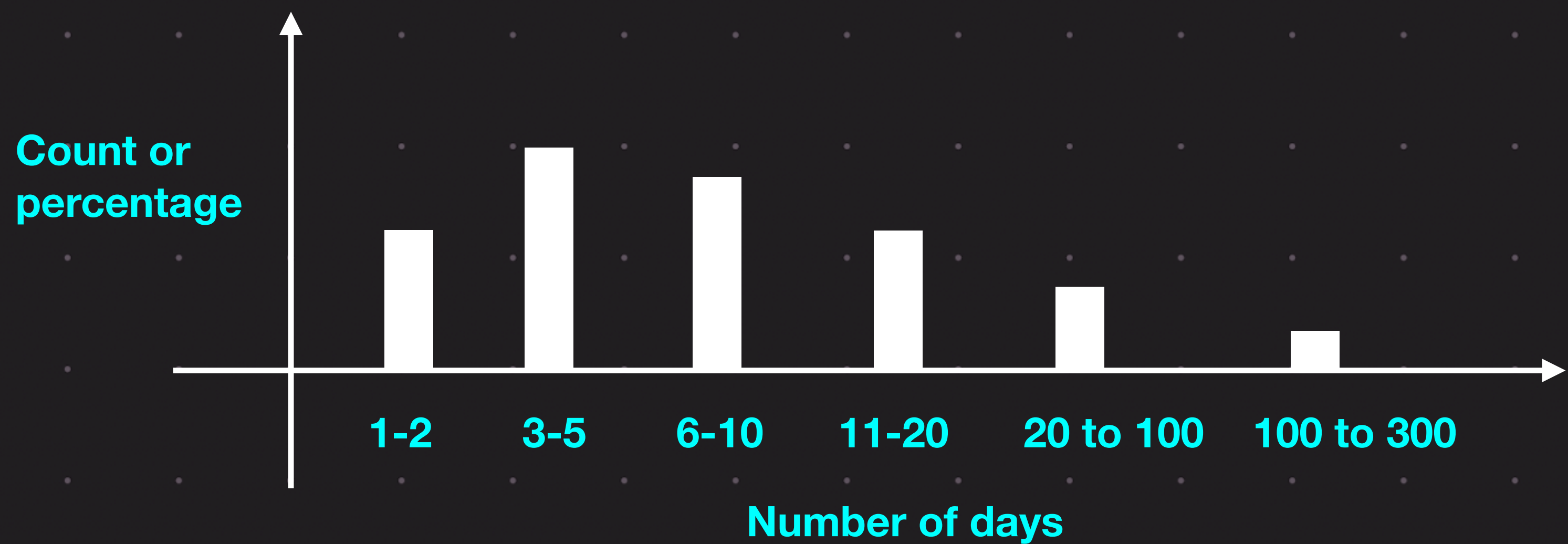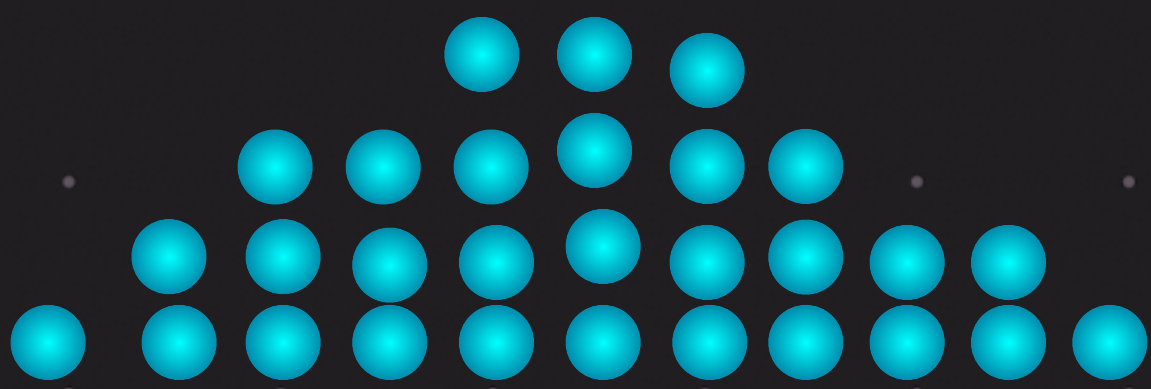
Log-Normal distribution

For the log normal, we use the parameters of the Gaussian: μ and σ

$$E[X] = e^{\mu + \sigma^2/2}$$

$$\text{Var}[X] = (e^{\sigma^2} - 1)\left(e^{2\mu + \sigma^2}\right)$$

Always confirm these equations from some source (eg: wiki)

Hours

X

σ

μ

Log of Hours

log(X)