Poisson Distribution

What is the Poisson Distribution for?

To predict the # of events that may occur in an interval.

Suppose you run an Internet shopping mall.

How many people will visit my shopping mall website per hour, day, week, etc.?

How many of them will actually purchase items from my mall in a week?

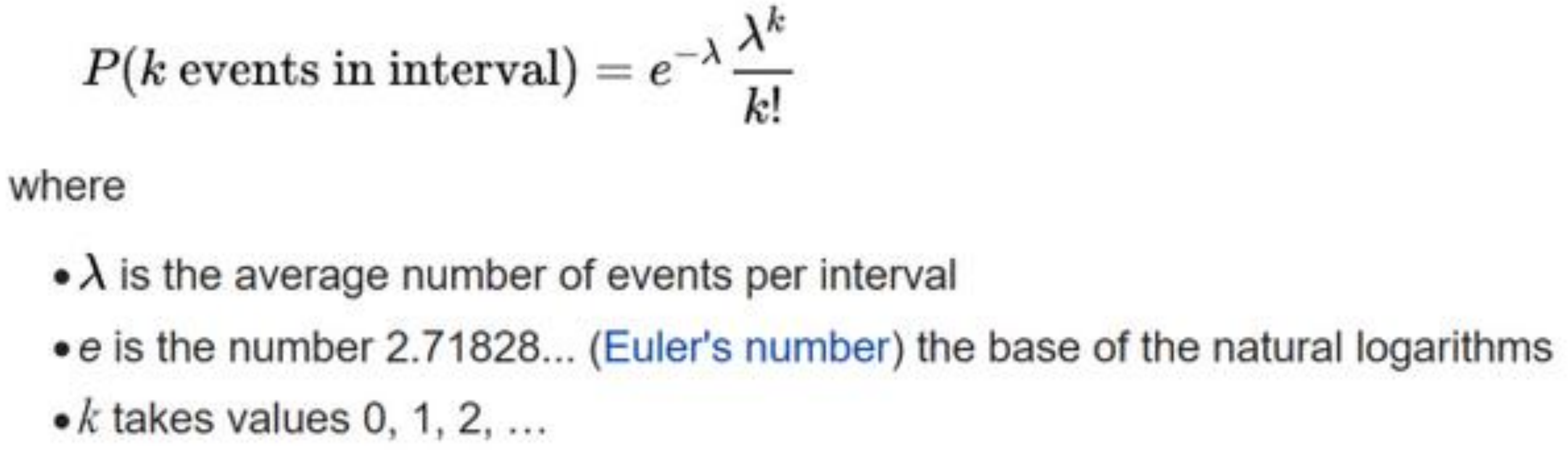
How many of them will return their purchased items within a month?

How many of those purchased customers will post review within a month?

Convenience of Poisson Distribution!

The only parameter of the Poisson distribution is the rate λ, the expected value of X.

The probability of observing *k* events in an interval is:



$$P(k \text{ events in interval}) = e^{-\lambda}\frac{\lambda^k}{k!}$$

where

- λ is the average number of events per interval
- *e* is the number 2.71828... (Euler's number) the base of the natural logarithms
- k takes values 0, 1, 2, …

Variance is also λ.

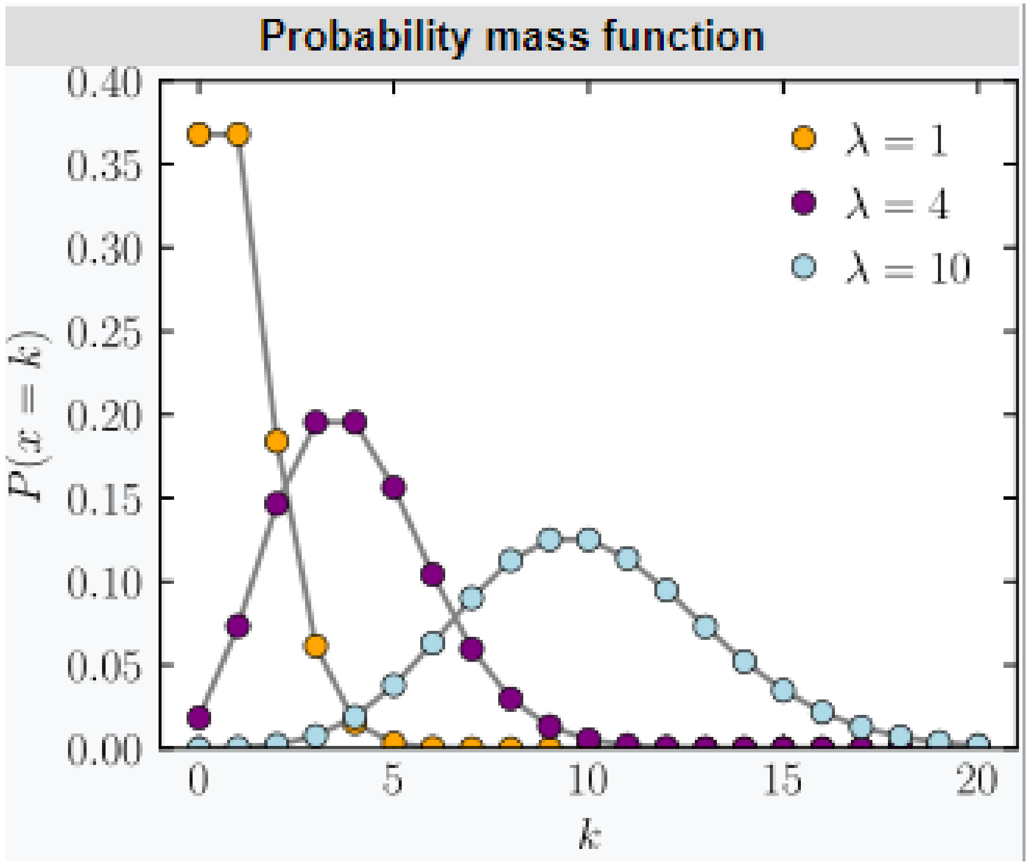
https://towardsdatascience.com/poisson-distribution-intuition-and-derivation-1059aeab90d

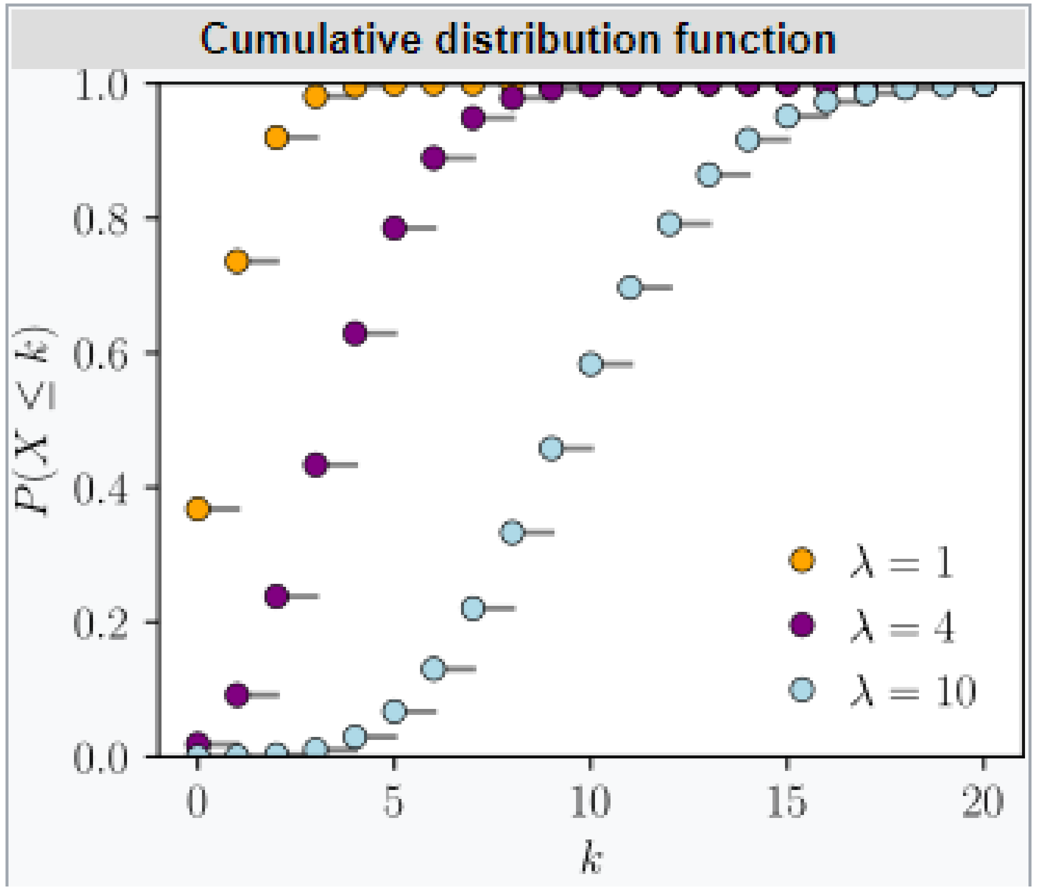
Poisson distribution typically is to model rare events.

The rate parameter λ modeling rare events can be any number -- it does not have to be a small number.

The Poisson Distribution is asymmetric — it is always skewed toward the right. Because it is inhibited by the zero occurrence barrier (there is no "minus" number of events) on the left and it is unlimited on the other side.

As λ becomes bigger, the graph looks more like a normal distribution.





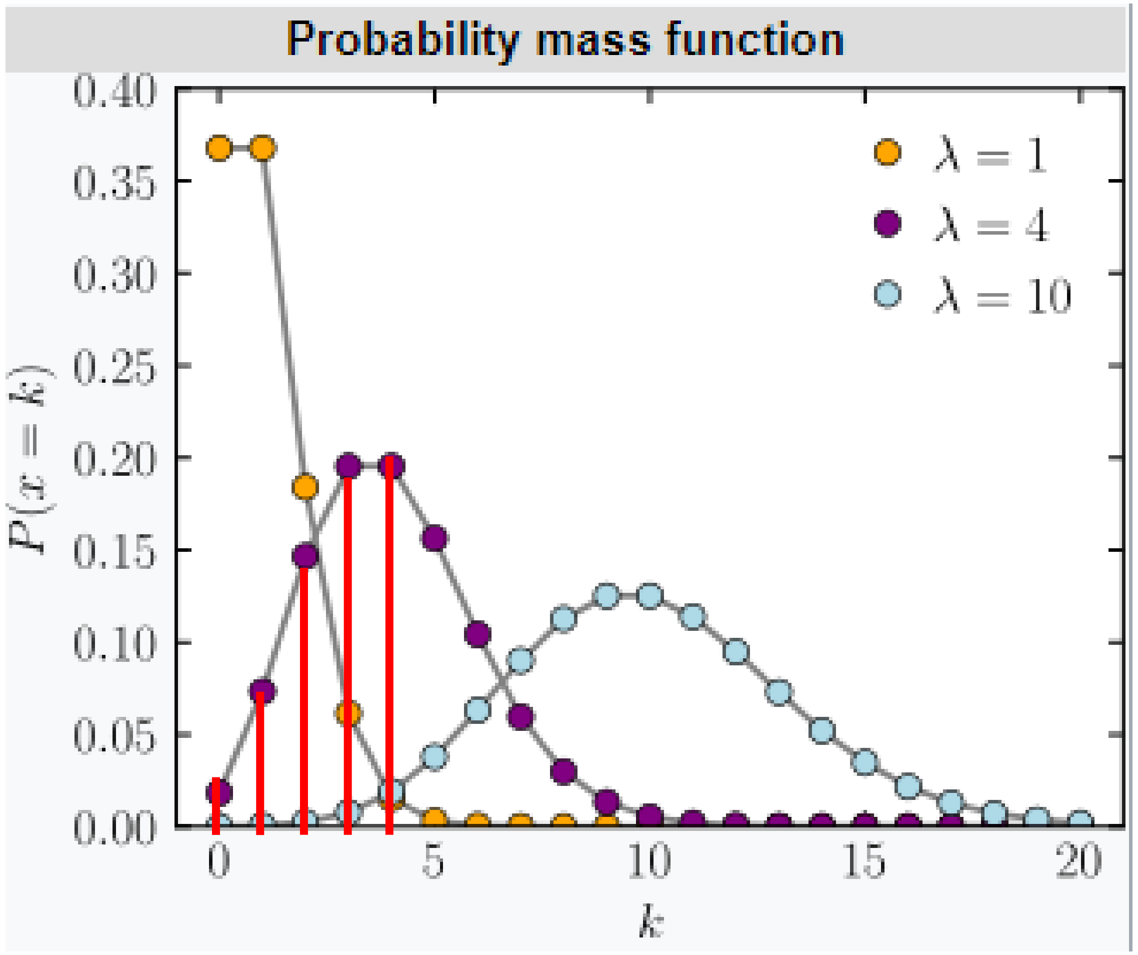
https://en.wikipedia.org/wiki/Poisson_distribution

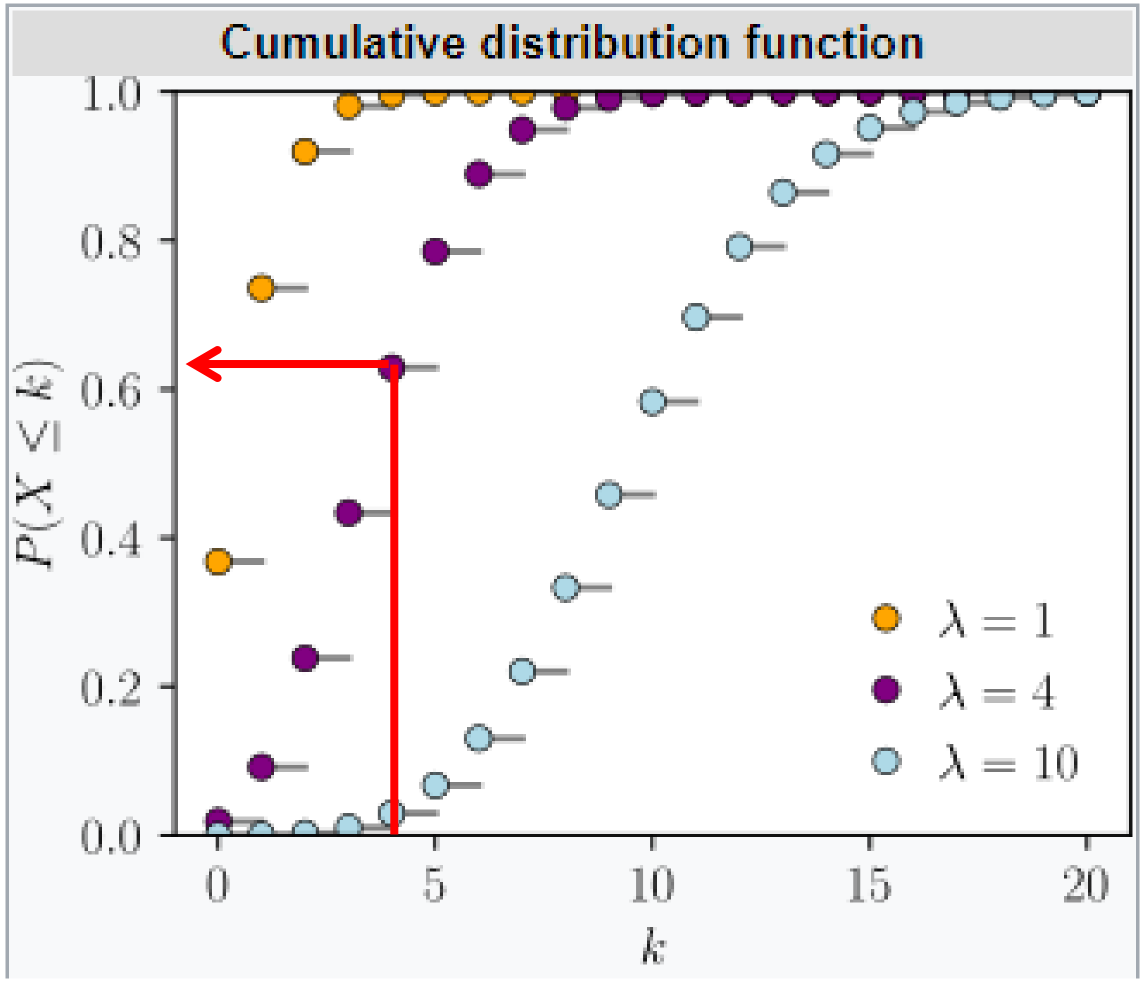
Poisson distribution typically is to model rare events.

Let λ = 4 be the number of houses a real estate agent Sue can sell per months.

What would be the probability that Sue can indeed sell four houses next month?

Show value!



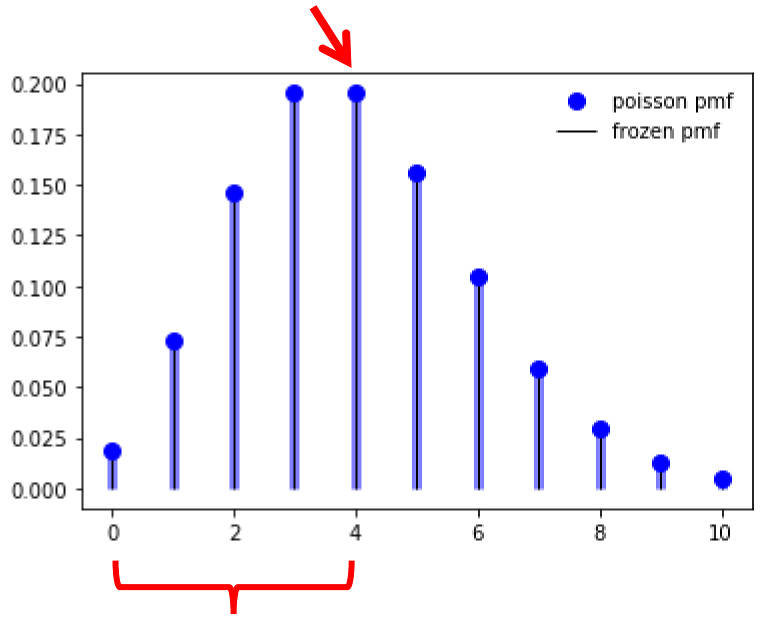


Assume $4,000 commission per house sold. What is the probability that Sue makes $16,000 next months?

Poisson Distribution

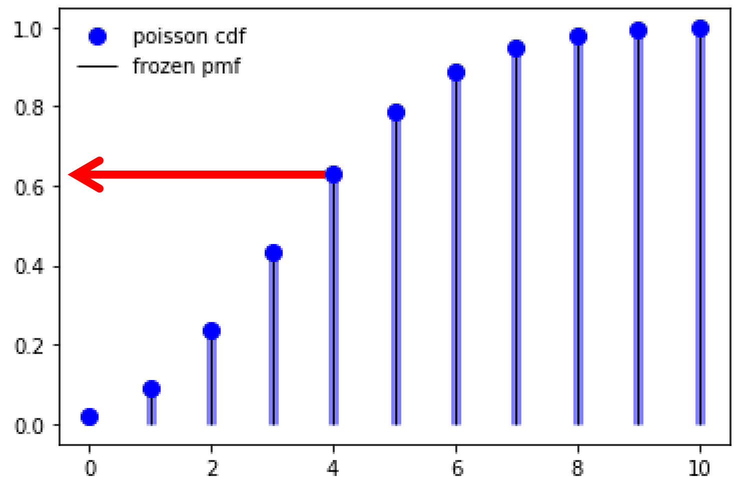
```
poisson.pmf(4,mu)
0.19536681481316454
```

```
import numpy as np
from scipy.stats import poisson
import matplotlib.pyplot as plt

fig, ax = plt.subplots(1, 1)

mu = 4 # mean - lambda; variance is also lambda.
mean, var, skew, kurt = poisson.stats(mu, moments='mvsk')
x = np.arange(poisson.ppf(0.001, mu),
              poisson.ppf(0.999, mu))

ax.plot(x, poisson.pmf(x, mu), 'bo', ms=8, label='poisson pmf')
ax.vlines(x, 0, poisson.pmf(x, mu), colors='b', lw=5, alpha=0.5)
rv = poisson(mu)
ax.vlines(x, 0, rv.pmf(x), colors='k', linestyles='-', lw=1,
                     label='frozen pmf')

ax.legend(loc='best', frameon=False)
plt.show()
```



```
poisson.cdf(4,mu)
0.6288369351798734
```

If you change pmf to cdf →



Gamma (Γ) Distribution

Clarifying two different parameterization sets — (k, θ) or (α, β)

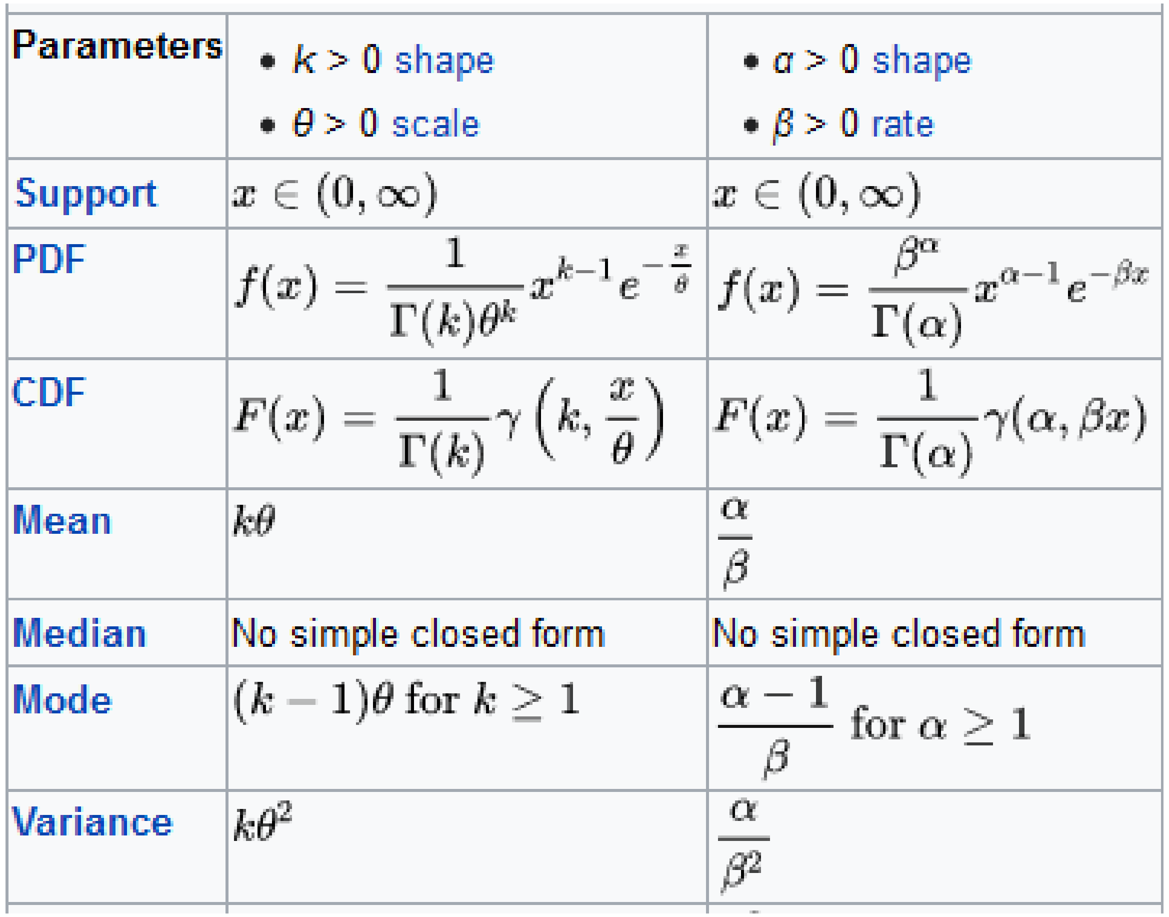
For (k, θ) parameterization: θ is a reciprocal of the event rate λ, which is the mean wait time (the average time between event arrivals).

$$\theta = 1/\lambda$$

For (α, β) parameterization: Substitute k (the # of events) with α and λ (the rate of events) with β. $\beta = \lambda$

https://towardsdatascience.com/gamma-distribution-intuition-derivation-and-examples-55f407423840

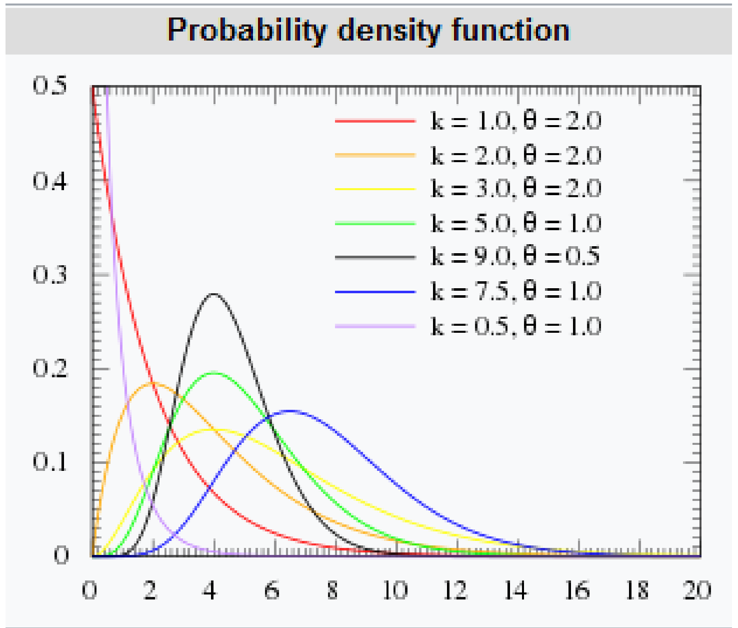
Gamma (Γ) Distribution

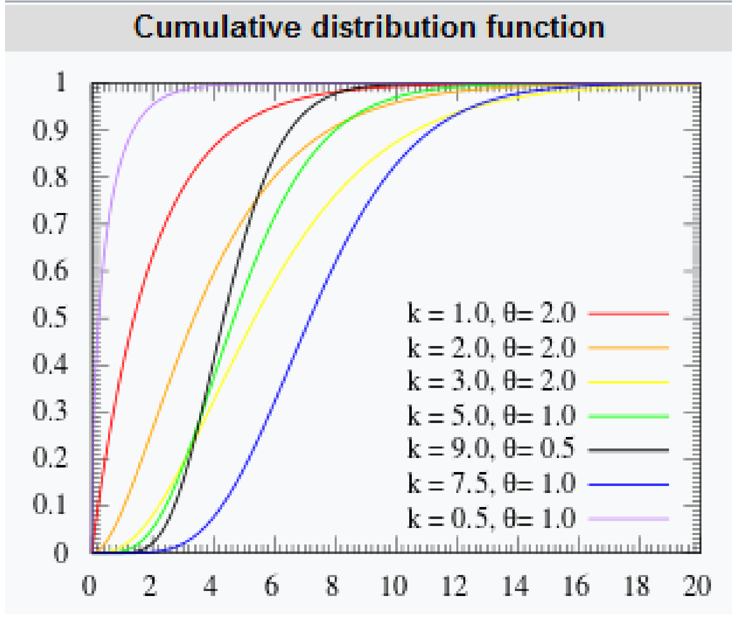


Parameters	• $k > 0$ shape • $\theta > 0$ scale	• $\alpha > 0$ shape • $\beta > 0$ rate
Support	$x \in (0, \infty)$	$x \in (0, \infty)$
PDF	$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$
CDF	$F(x) = \frac{1}{\Gamma(k)} \gamma\left(k, \frac{x}{\theta}\right)$	$F(x) = \frac{1}{\Gamma(\alpha)} \gamma(\alpha, \beta x)$
Mean	$k\theta$	$\frac{\alpha}{\beta}$
Median	No simple closed form	No simple closed form
Mode	$(k-1)\theta$ for $k \geq 1$	$\frac{\alpha - 1}{\beta}$ for $\alpha \geq 1$
Variance	$k\theta^2$	$\frac{\alpha}{\beta^2}$

$\alpha = k$

$\beta = 1/\theta$



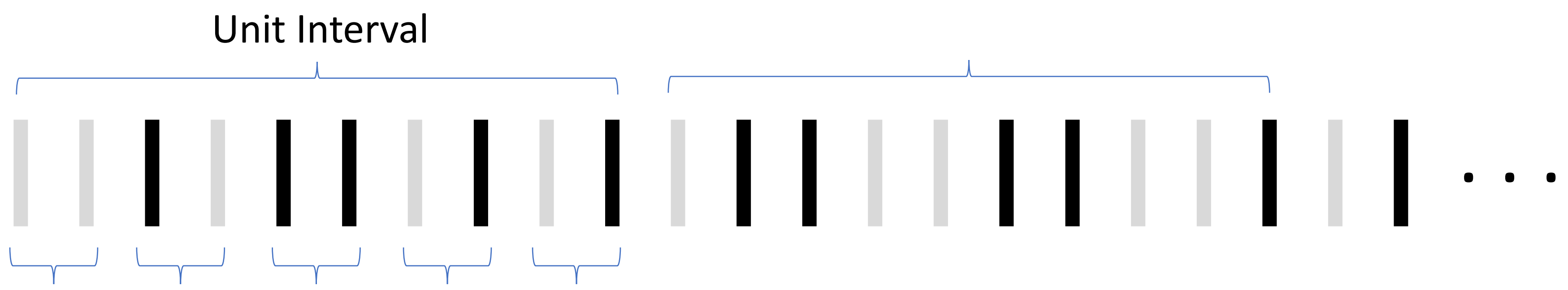


https://en.wikipedia.org/wiki/Gamma_distribution

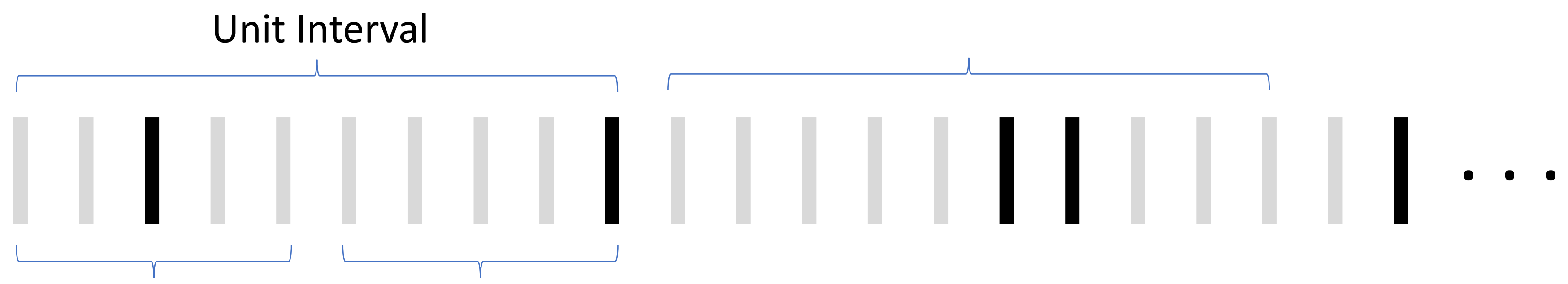
# of event occurrences & Rate of event occurrence

What is this Gamma (Γ) Distribution for?

To predict the wait time until future events, specifically until the “k-th” event occurs.



λ = 5 → reciprocal θ = 1/5 = 0.2 (every 0.2 interval the event occurs).



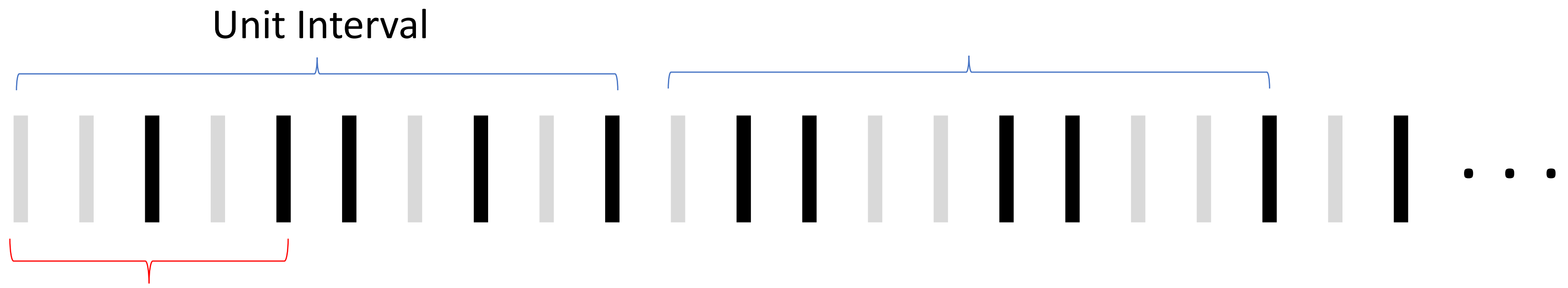
λ = 2 → reciprocal θ = 1/2 = 0.5 (every 0.5 interval the event occurs).

Time for k-th event to occur

To predict the wait time until future events, specifically until the "k-th" event occurs.
Assume Sue and John are two real estate agents. Sue is selling houses faster than John. Assume 10 showings per month in each case with "**Sell**" and "**No-sell**".

Sue

λ = 5, or
$1/\lambda$ = 0.2



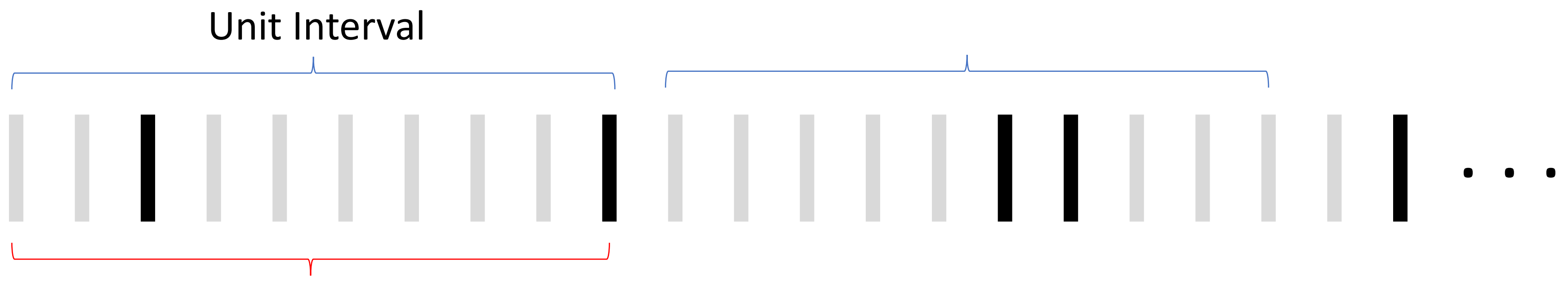
k = 1 → After 3 trials.
k = 2 → After 5 trials.

k = 1 → **After ? trials.**
k = 2 → **After ? trials.**

On average, how many trials for Sue to sell 2 houses? Translate that trials into portion of unit time?

John

λ = 2, or
$1/\lambda$ = 0.5



k = 1 → After 3 trials.
k = 2 → After 10 trials.

k = 1 → **After ? trials.**
k = 2 → **After ? trials.**

On average, how many trials for Sue to sell 2 houses? Translate that trials into portion of unit time?

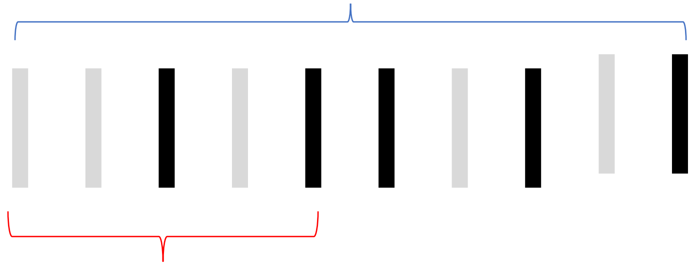
Time for k-th event to occur

How long should Sue and John wait, respectively, to sell their respective two houses?

Sue

λ = 5, or
$1/\lambda$ = 0.2

Unit Interval

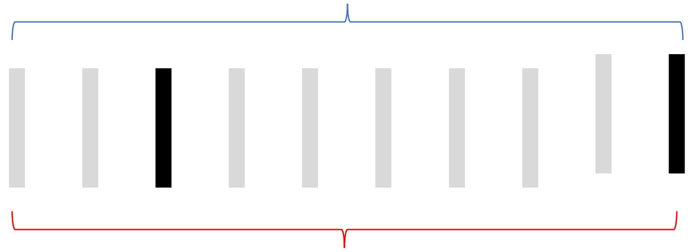


k = 1 → After 3 trials.
k = 2 → After 5 trials.

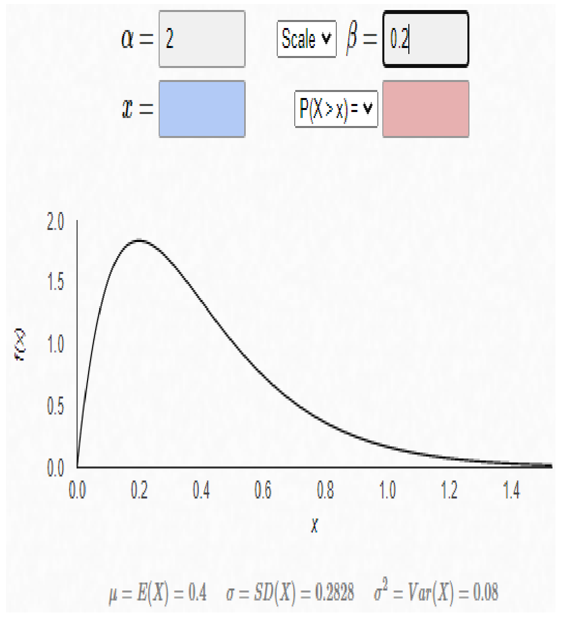
John

λ = 2, or
$1/\lambda$ = 0.5

Unit Interval



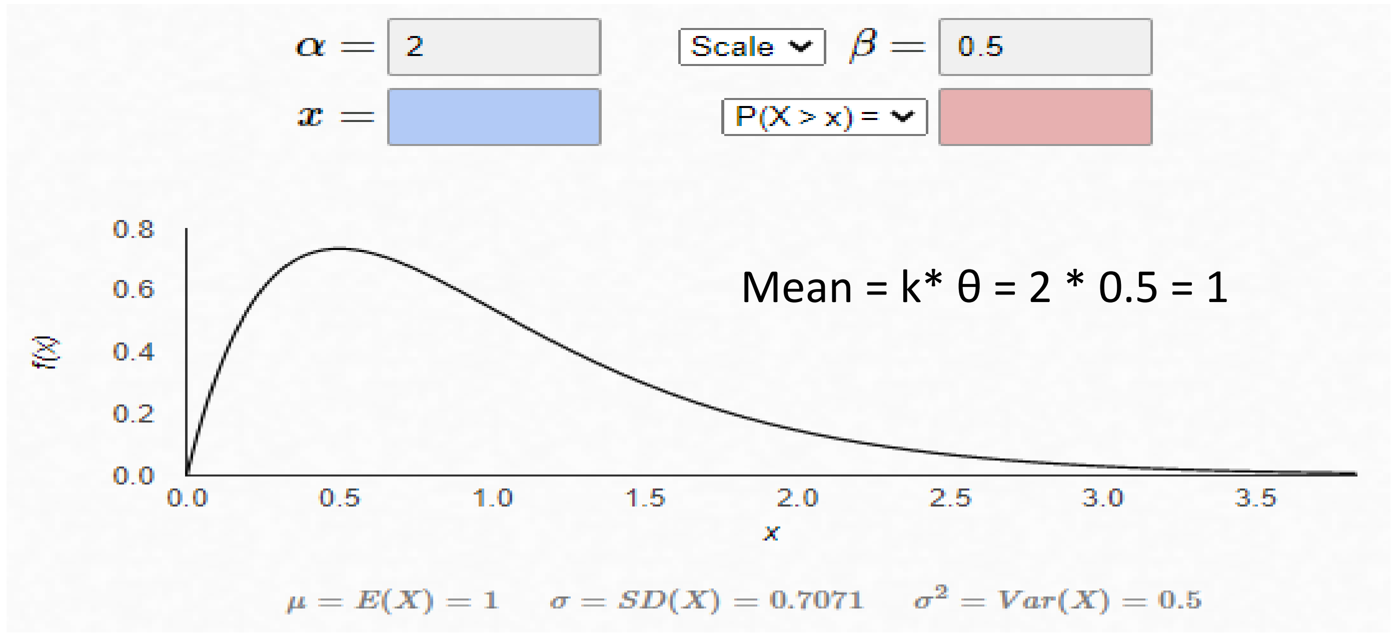
k = 1 → After 3 trials.
k = 2 → After 10 trials.



α = 2   Scale β = 0.2
x =   P(X>x)=

$\mu = E(X) = 0.4 \quad \sigma = SD(X) = 0.2828 \quad \sigma^2 = Var(X) = 0.08$

Mean = k* θ = 2 * 0.2 = 0.4



α = 2   Scale β = 0.5
x =   P(X > x) =

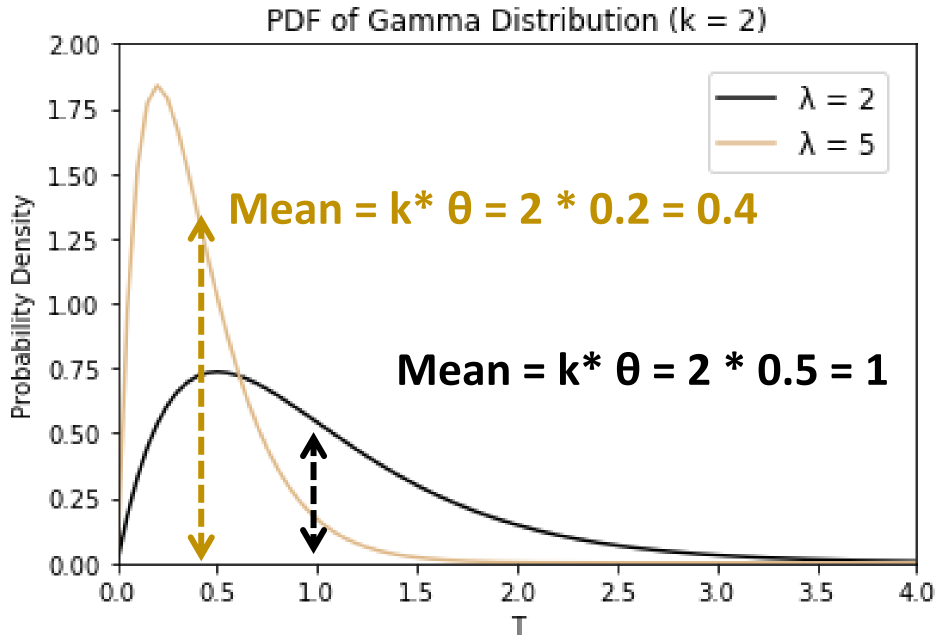
$\mu = E(X) = 1 \quad \sigma = SD(X) = 0.7071 \quad \sigma^2 = Var(X) = 0.5$

Mean = k* θ = 2 * 0.5 = 1

https://homepage.divms.uiowa.edu/~mbognar/applets/gamma.html

```
import numpy as np
from scipy.stats import gamma
import matplotlib.pyplot as plt

def plot_gamma_lambda():
    """
    k : the number of events waiting to occur.
    λ : the rate of events happening following Poisson D.
    """
    a = 2  # k = 2
    x = np.linspace(0, 50, 1000)
    lambda_ = 2
    mean, var, skew, kurt = gamma.stats(a, scale=1/lambda_, moments='mvsk')
    y1 = gamma.pdf(x, a, scale=1/lambda_)
    lambda_ = 5
    mean, var, skew, kurt = gamma.stats(a, scale=1/lambda_, moments='mvsk')
    y2 = gamma.pdf(x, a, scale=1/lambda_)
        plt.title("PDF of Gamma Distribution (k = 2)")
    plt.xlabel("T")
    plt.ylabel("Probability Density")
    plt.plot(x, y1, label="λ = 2", color='black')
    plt.plot(x, y2, label="λ = 5", color='burlywood')
    plt.legend(bbox_to_anchor=(1, 1), loc='upper right', borderaxespad=1, fontsize=12)
    plt.ylim([0, 2.0])
    plt.xlim([0, 4.0])

plot_gamma_lambda()
```

Sue:
Wait 0.4 month.



John:
Wait 1 month.

Gamma (Γ) Distribution

Gamma distribution is a distribution that arises naturally in processes for which **the waiting times between events are relevant**. It can be thought of as **"a waiting time"** between Poisson distributed events.

Example Suppose you are fishing and you expect to get a fish once every 1/2 hour. Compute the probability that you will have to **wait between 2 to 4 hours** before you catch **4 fish.** **Two fishes per hour!**

One fish every 1/2 hour → $\theta = 1 / 0.5 = 2$ fish every hour on average.

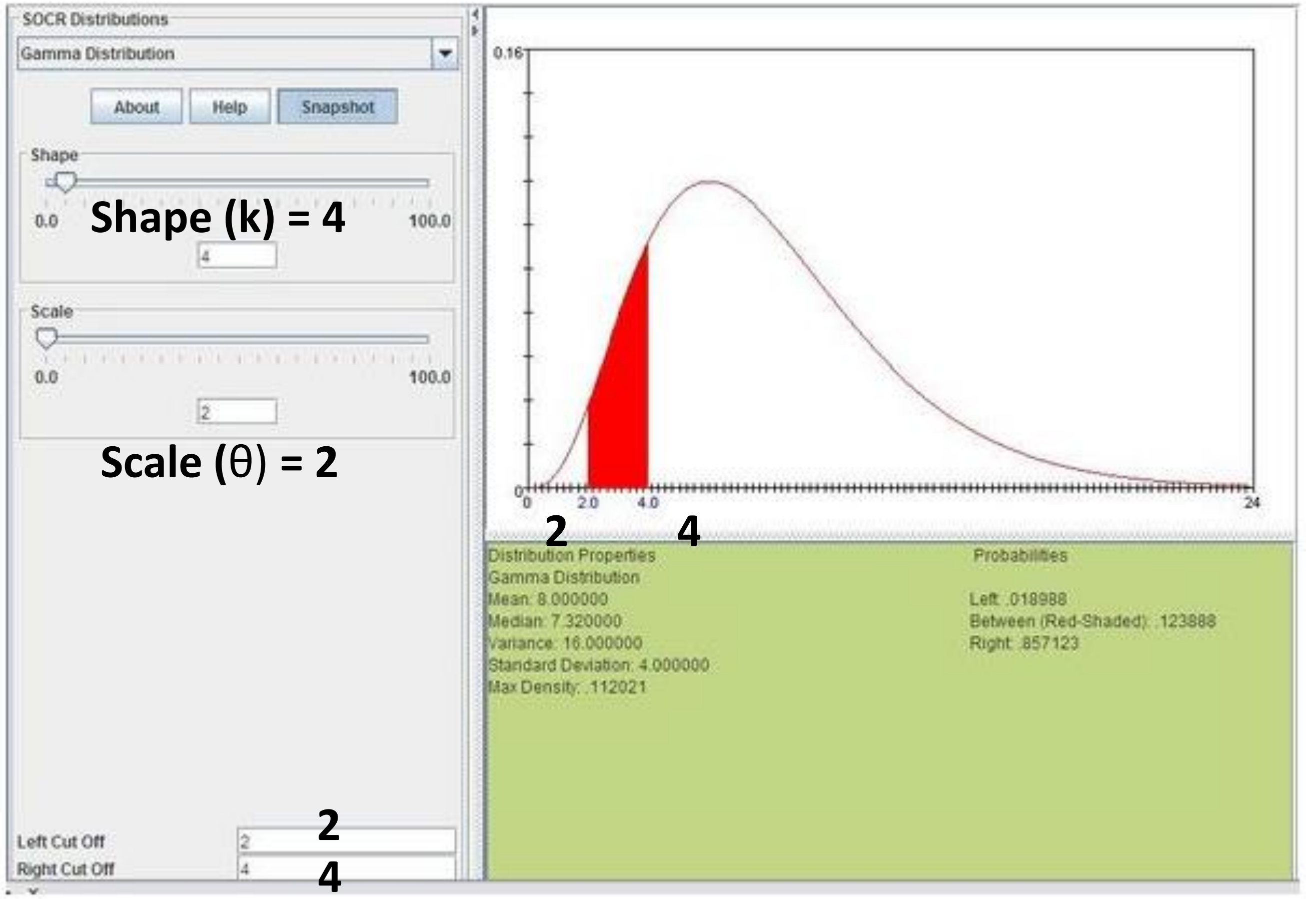
Using $\theta = 2$ and $k = 4$, we can compute this as follows:

$$P(2 \leq X \leq 4) = \sum_{x=2}^{4} \frac{x^{4-1} e^{-x/2}}{\Gamma(4) 2^4} = 0.12388$$

PDF	$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$

$P(X>2) = 0.98101$ $P(X>4) = 0.85712$

$P(4>X>2) = 0.98101 - 0.85712 = 0.12389$



Shape
0.0 100.0
4
Shape (k) = 4

Scale
0.0 100.0
2
Scale (θ) = 2

2 4

Distribution Properties
Gamma Distribution
Mean: 8.000000
Median: 7.320000
Variance: 16.000000
Standard Deviation: 4.000000
Max Density: .112021

Probabilities
Left: .018988
Between (Red-Shaded): .123888
Right: .857123

Left Cut Off 2 2
Right Cut Off 4 4

http://wiki.stat.ucla.edu/socr/index.php/AP_Statistics_Curriculum_2007_Gamma