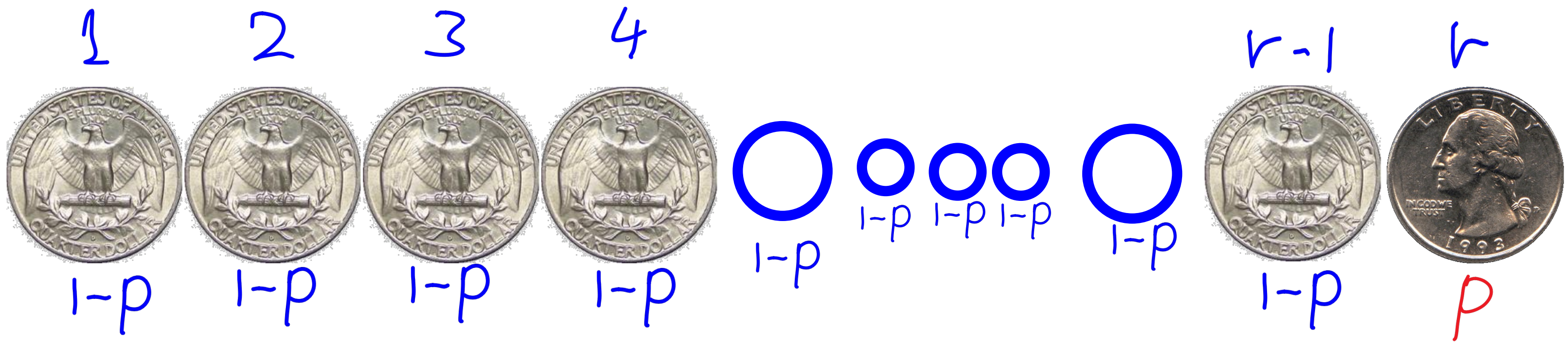
The Geometric Exponential and Laplace distributions

Number of coin flips until first heads

- Suppose we have a biased coin whose probability of <u>*heads*</u> is $0 < p < 1$
- We flip the coin until we get <u>*heads*</u>
- What is the expected number of coin flips until we get the first <u>*heads*</u>?



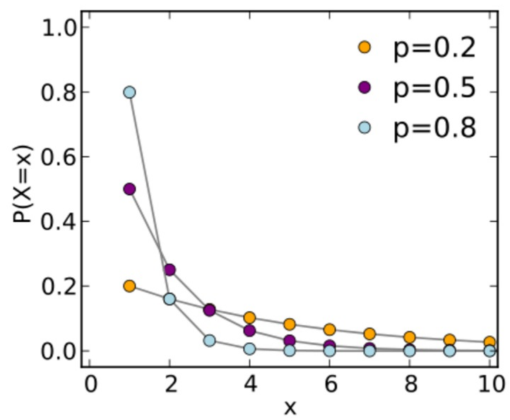
The geometric distribution

- Let $X_1, X_2, \ldots$ be IID Binary RV such that $P(X_i = 1) = p$
- The prob. that the 1st coin flip is the first heads is p
- The prob. that the 2nd coin flip is the first heads is $(1-p)p$
- The prob. that the 3rd coin flip is the first heads is $(1-p)^2 p$
- ….
- The prob. that the rth coin flip is the first heads is $(1-p)^{r-1} p$
- This is the geometric distribution: $P(R = r) = (1-p)^{r-1} p$

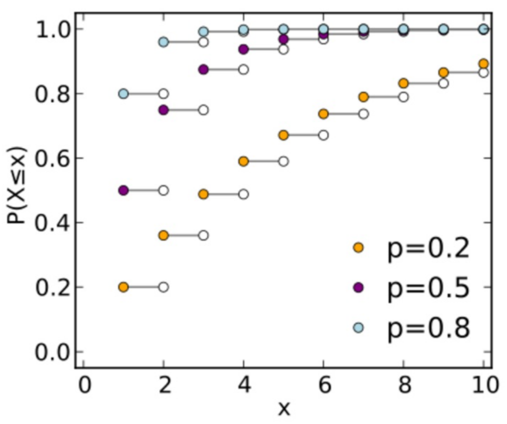
The geometric distribution

Parameters	$0 < p \leq 1$ success probability (real)
Support	k trials where $k \in \{1, 2, 3, \dots\}$
Probability mass function (pmf)	$(1-p)^{k-1}\,p$
CDF	$1-(1-p)^k$

Probability mass function



Cumulative distribution function



$$E[x] = \frac{1}{p} \qquad Var[x] = \frac{1-p}{p^2}$$

Expected value: Solution 1

- The expected value of R is:

$$E[R] = \sum_{r=1}^{\infty} r(1-p)^{r-1}p \;\; = \frac{p}{1-p}\sum_{r=1}^{\infty} r(1-p)^r$$

- We can use the formula for power series:

$$\sum_{i=0}^{\infty} ia^i = \frac{a}{(1-a)^2}$$

- Substituting $i = r,\;\; a = (1-p)$ we get:

$$E[R] = \frac{p}{1-p}\sum_{r=1}^{\infty} r(1-p)^r \;\; = \frac{p}{1-p}\frac{(1-p)}{p^2} = \frac{1}{p}$$

- This makes intuitive sense: the expected number of times we need to flip a coin with probability of heads $p = \frac{1}{10}$ until we get the first heads is $\frac{1}{p} = 10.$

Expected value: Solution 2

- We can find the expected number of coins using a simple recursive formula.
- $E[R]$ is the expected number of coin flips until the first heads.
- Consider the first coin flip:
  - With probability p we get heads – we used exactly one coin flip.
  - With probability $1-p$ we get tails – we used one coin flip and we need an expected number of $E[R]$ additional coin flips to get the first head.
- This can be written as follows:

$$E[R] = p \times 1 + (1-p) \times (1 + E[R])$$

- Which simplifies to $E[R] - (1-p) \times E[R] = p + (1-p) = 1$
- $E[R] = \frac{1}{p}$

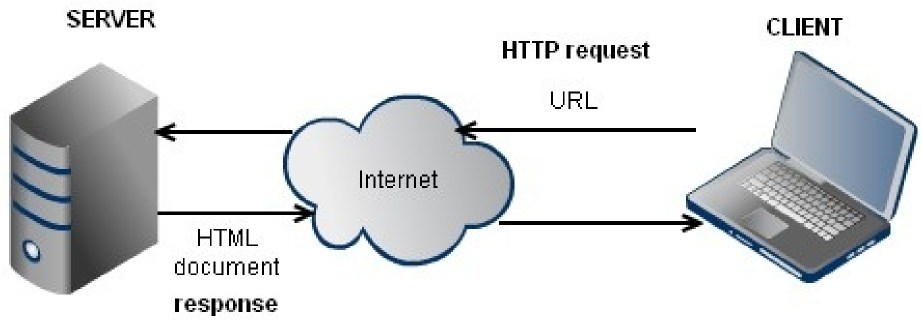
The bet doubling Strategy

- always bet on heads
- Stop after first heads. (1\$)
- Bet \$1, \$2, \$4 ..... $\$2^t$
- Loss: \$1, \$3, \$7, ... $\$(2^{t+1}-1)$
- Suppose we start with \$1023
- Game can end in 2 ways:
  - Win +\$1 : Complement:
  - Lose -\$1023 : 10 tails in a row - prob $= 2^{-10}$

$$E(\text{Gain}) = \frac{-\$1023}{1024} + \left(1 - \frac{1}{1024}\right) \times \$1 = \$0$$

Theorem: a stopping rule that depends only on the past cannot change the expected value of a game of chance

The Exponential Distribution

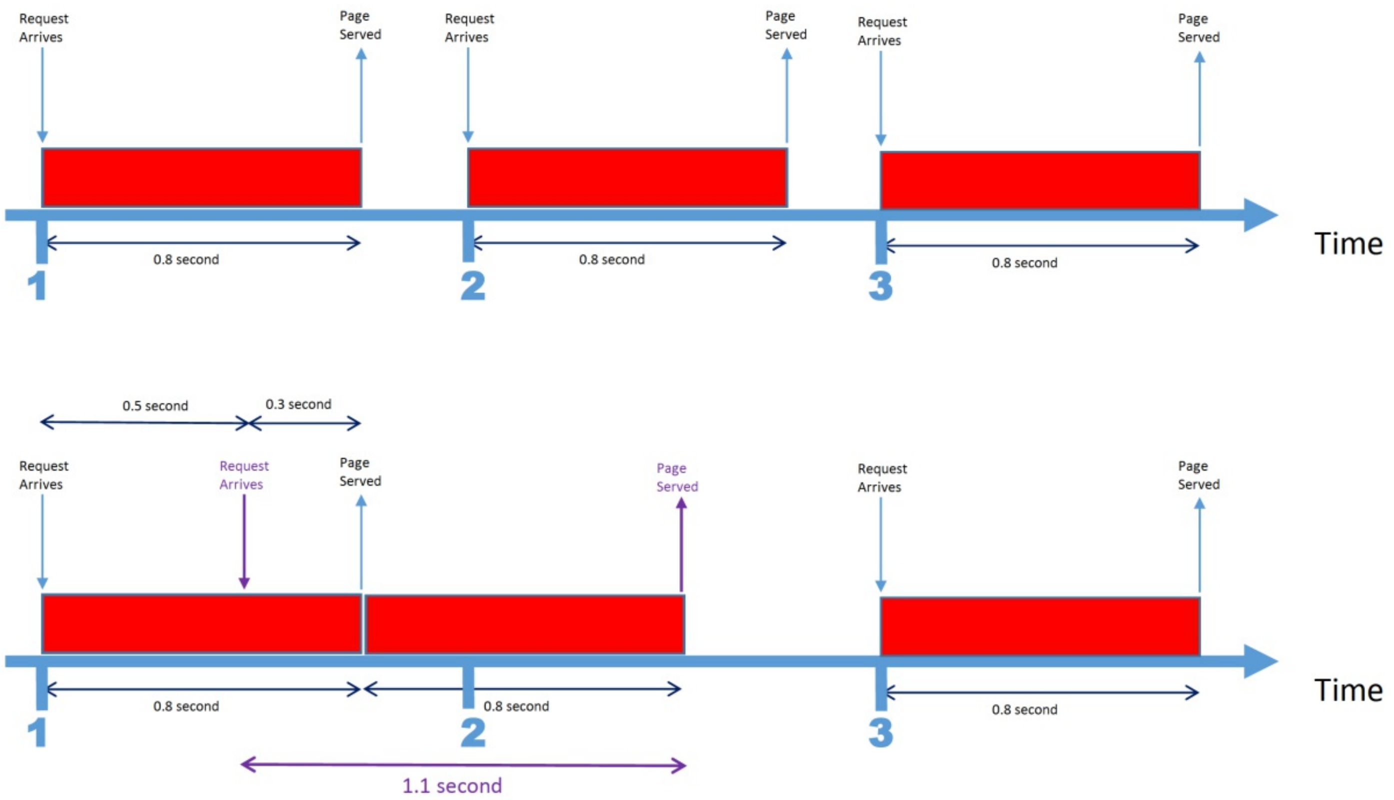


Suppose we are hosting a commercial web site on our web server.

* It takes our server 0.8sec to serve a request.

* We have a "service level agreement" (SLA) contract with our clients (the web site owners) in which we promise to serve 99% of the requests within 1 second.

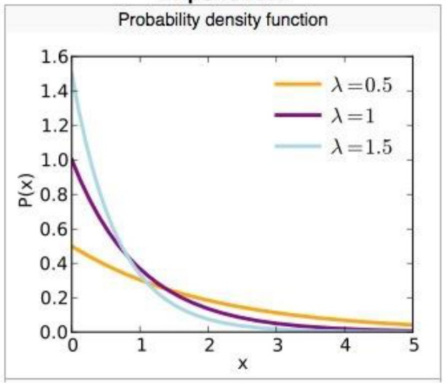
* What is the rate of requests we can serve and adhere to the SLA?

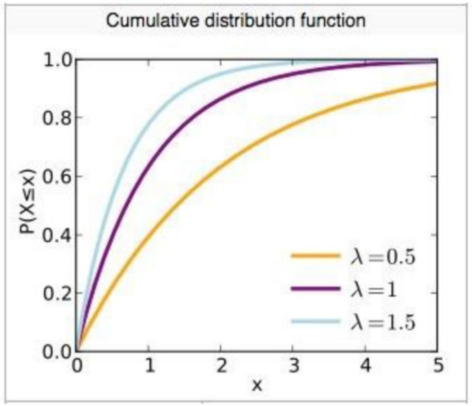


If gap between consecutive requests Is less than 0.6 sec then
The second request will take more than 1 second to complete.

Exponential distribution

Exponential





Parameters	$\lambda > 0$ rate, or inverse scale
Support	$x \in [0, \infty)$
PDF	$\lambda e^{-\lambda x}$
CDF	$1 - e^{-\lambda x}$

$$E[x] = \frac{1}{\lambda} \qquad Var[x] = \frac{1}{\lambda^2}$$

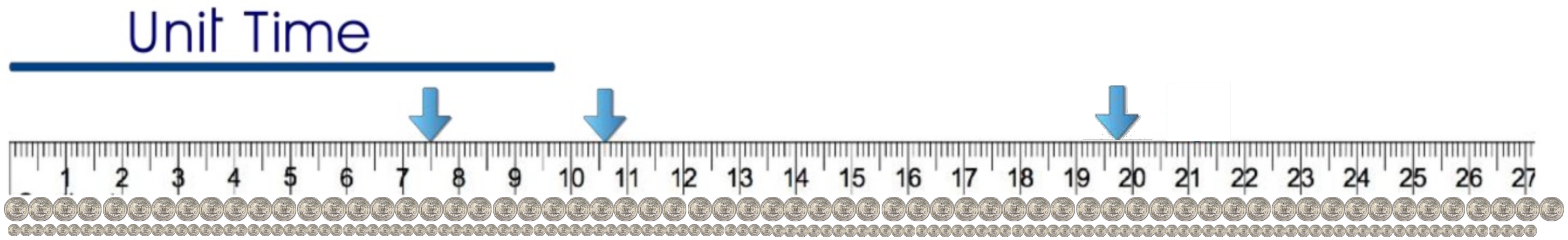
What is the maximal rate we can tolerate?

- The probability that the time gap between two consecutive requests is less than 0.6 Sec should be smaller than 1/100.
- The CDF of the exponential distribution is
$$F(t) = 1 - e^{-\lambda t}$$
- We require $F(0.6) \le \frac{1}{100}$ therefor $1 - e^{-0.6\lambda} \le \frac{1}{100}$
- Or $e^{-0.6\lambda} \ge \frac{99}{100} \Rightarrow -0.6\lambda \ge \ln\frac{99}{100} \Rightarrow \lambda \le \frac{\ln\frac{100}{99}}{0.6} = 0.0167$
- In words, the rate of requests must be less than 0.0167, or about one request per minute!
- What does this mean in practice?
  - Use a computer fast enough that it can serve 3-4 requests within a second.
  - Use queues with more than one place.
  - Use Queuing theory to compute the distribution of time for queue+process.
- **In order to serve random requests within time constraints the computer has to be idle a significant fraction of the time!**

The exponential distribution

- Consider the following situation: a server receives requests at random at an average rate of λ requests per second.
- What is the distribution of the time lapse between consecutive requests?
- This distribution smooth over the reals, so it has to be a density distribution. (there are no particular time gaps that have non zero probability)
- Suppose we divide each second into n equal-length segments.
- The probability that there will be a request in a particular segments is $\mathrm{p} = \lambda/n$.
- We got a geometric distribution where the probability heads = probability of a request in any particular segment is p

Constant Rate: Discretizing the time line



Fix:

1. The rate of events: λ
2. Unit time: $t = 1$

Scale:

1. The number of bins: $n \to \infty$
2. The probability that a particular event occurs within a particular bin: $p \to 0$

$$\lambda \doteq E(\#\text{events in unit time}) = np \quad \Rightarrow \quad p = \frac{\lambda}{n}$$

$n \to \infty$

$p \to 0$

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

The exponential distribution

- The expected number of segments until the first request is
$$\frac{1}{p} = \frac{n}{\lambda}$$
- The time length of each segment is $1/n$ second. Therefor the expected amount of time between requests is $1/\lambda$ seconds.
- This makes intuitive sense: if the rate of packets is 10 per second then the expected time between consecutive packets is 1/10 sec.
- Note: the expected time does not depend on n the number of segments in a second.
- To arrive at the density function of the distribution we let the number of segments go to infinity: $n \to \infty$

The exponential distribution

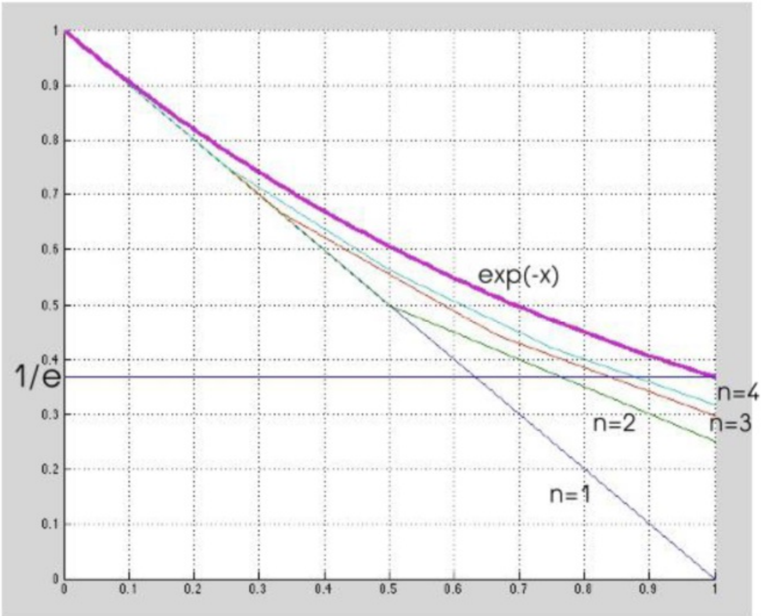
- Suppose we have n segments per second.
- What is the probability that the first packet will arrive after x seconds (in the xn'th segment)?

$$P(R = xn) = (1-p)^{xn-1}p = \left(1 - \frac{\lambda}{n}\right)^{xn-1} \frac{\lambda}{n} = \left(1 - \frac{\lambda}{n}\right)^{xn} \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right)^{-1}$$

- Recall that the density is defined as the limit of the ratio between the probability and the length of the segment, so we get

$$f(a) = \lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^{xn} \lambda\left(1 - \frac{\lambda}{n}\right)^{-1} = \lambda \lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^{xn} = \lambda e^{-\lambda x}$$

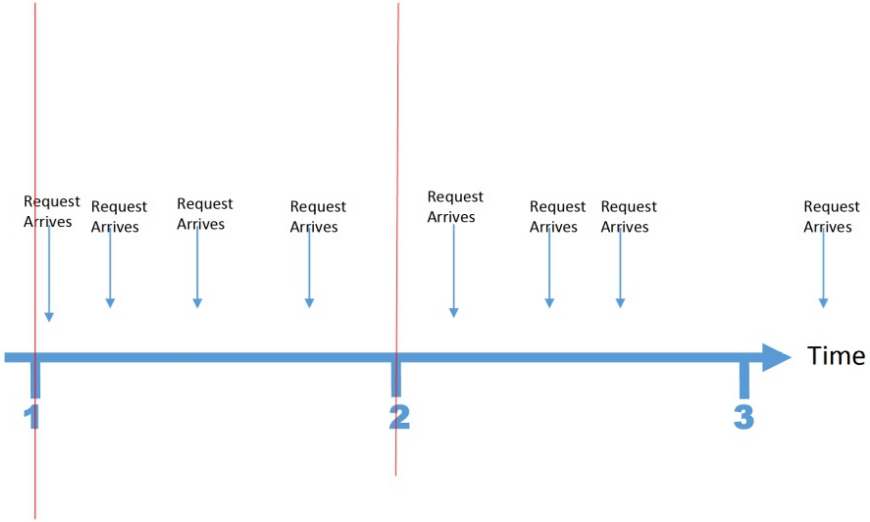
$$e \doteq \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n, \quad e^{-1} \doteq \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n$$



Compound interest on a loan

$$e^a = \lim_{n\to\infty}\left(1+\frac{a}{n}\right)^n$$

The Poisson distribution

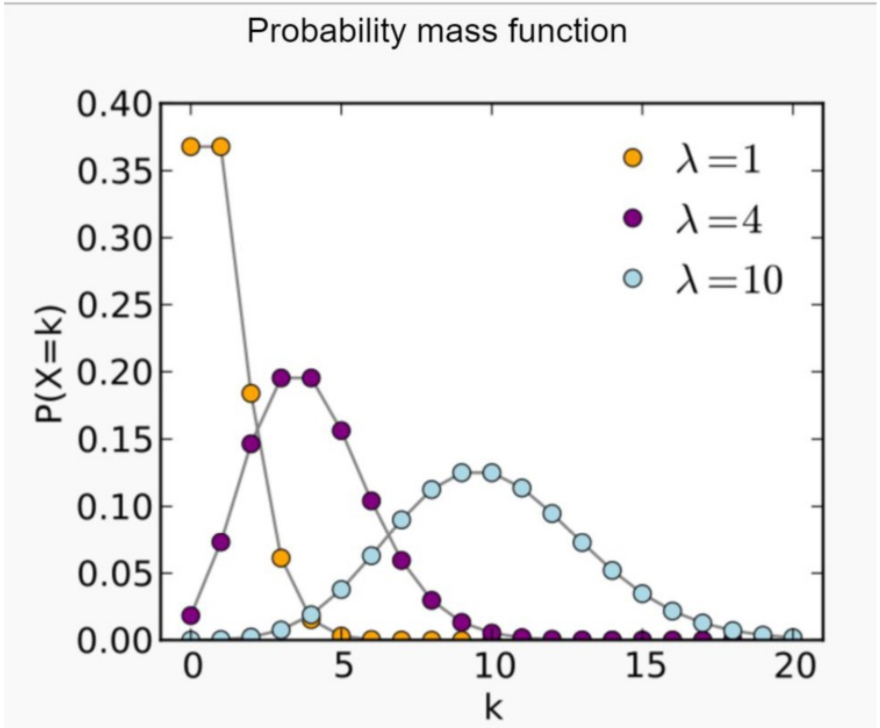

Request Arrives
Request Arrives
Request Arrives
Request Arrives
Request Arrives
Request Arrives
Request Arrives
Request Arrives
Time
1
2
3

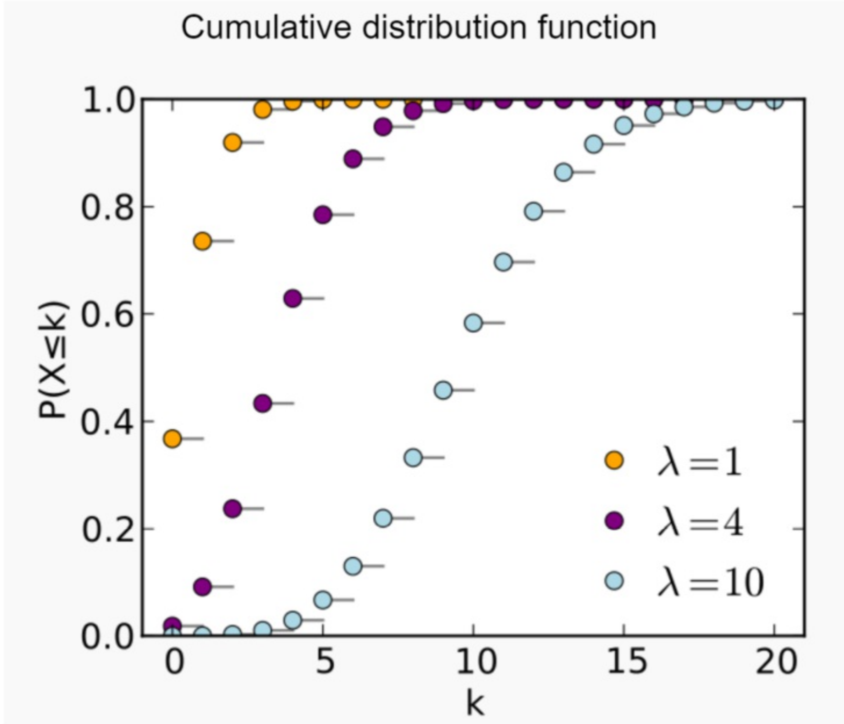
Counting requests

- The exponential distribution characterizes the time gap between consecutive requests.
- Consider a slightly different question: if the request rate is $\lambda = 100$ requests per second. What is the probability of receiving 120 requests during a particular second?
- The answer to this question is given by the Poisson distribution:
- The probability that k events occur within a time segment in which the expected number of events is λ is:

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

Poisson





Parameters	$\lambda > 0$ (real)
Support	$k \in \mathbb{N}$
pmf	$\frac{\lambda^k e^{-\lambda}}{k!}$
CDF	$e^{-\lambda} \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}$

$$E[k] = Var[k] = \lambda$$

Suppose requests arrive at our server independently at an average rate of 100 per second.

1. What is the probability that K requests arrive during a period of T seconds?

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

Suppose requests arrive at our server independently at an average rate of 100 per second.

2. What is the probability that the time gap beween two consecutive requests is larger than t seconds?

$$CDF(x) = 1 - e^{-\lambda x}$$

Suppose requests arrive at our server independently at an average rate of 100 per second.

3. Suppose our server consists of 100 independent cores, what is the probability that a core would be assigned l requests during a particular 1 second interval?

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