The binomial distribution when the number of flips is very large.

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

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

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

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?

The binomial distribution

The probability that when a coin, Pr(heads)=p, is flipped n times, the number of heads is k.

$X_1, X_2, \ldots, X_n$ are IID binary random variables

$\Pr(X_i = 1) = p.$

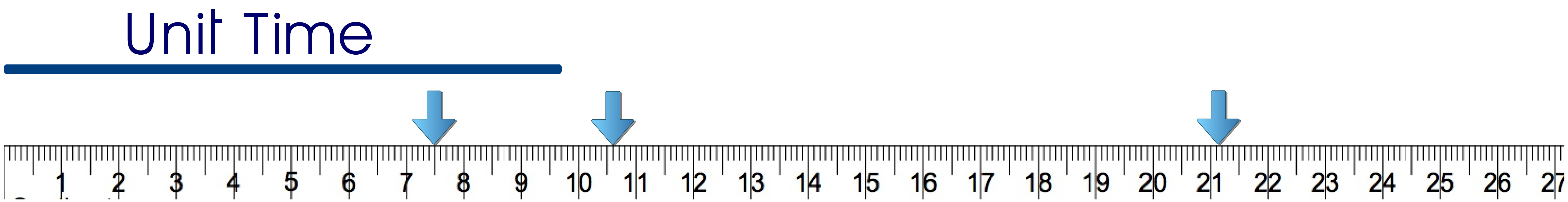
$$\Pr\left(\sum_{i=1}^{n} X_i = k\right) = b(n,p,k) \doteq \left(\begin{array}{c} n \\ k \end{array}\right) p^k (1-p)^{n-k}$$

We will consider two limits for $n \to \infty$:

1. Constant rate: $p = \dfrac{\lambda}{n}$

2. Constant probability: p is a constant.

Constant Rate: Discretizing the time line



Fix:

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

Scale:

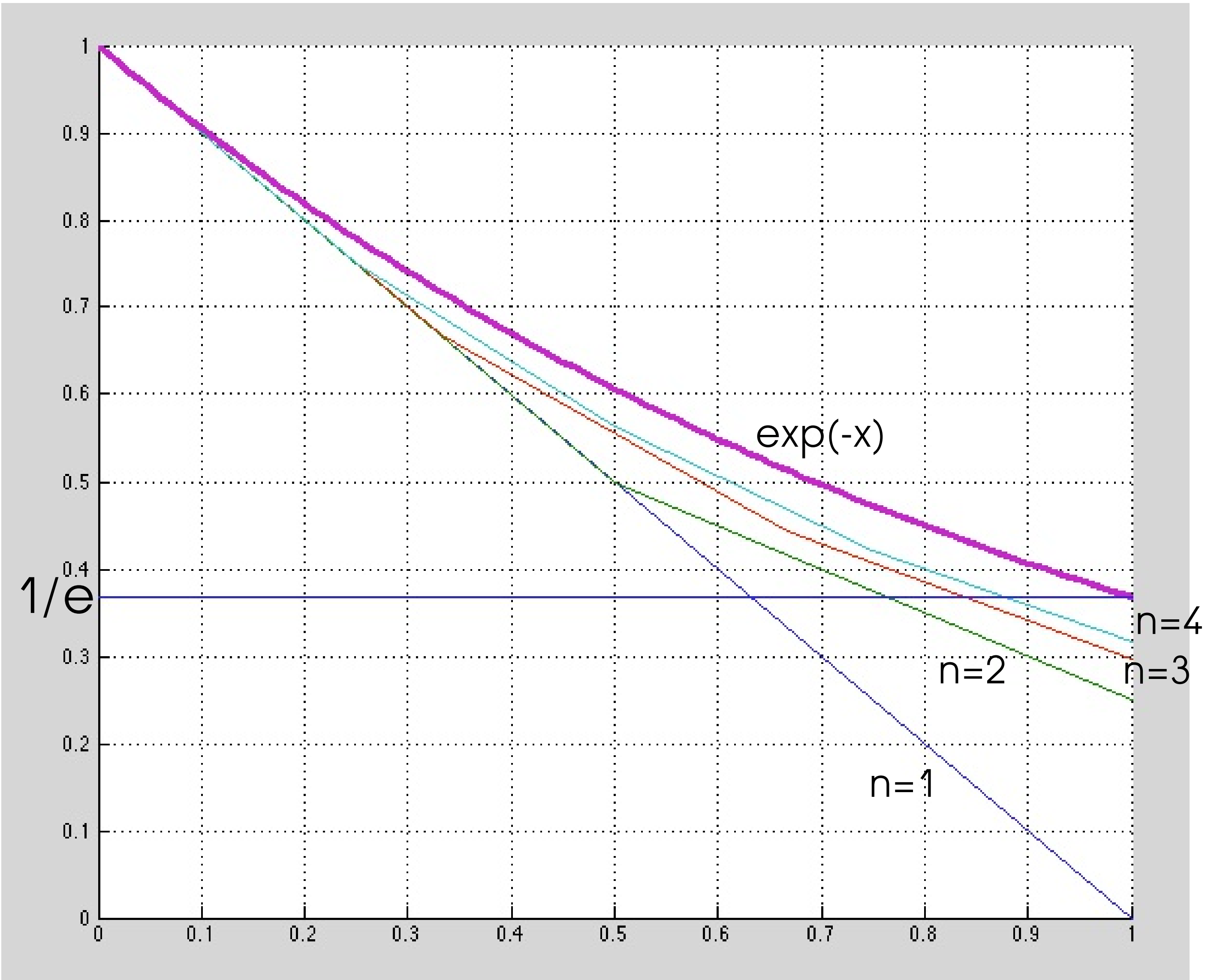
1. The number of bins: $n \to \infty$
2. The probability that a particular event occurs within a particular bin: $p \to 0$

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

What is b(n,p,0) ?

$$b(n,p,0) = (1-p)^n = \left(1-\frac{\lambda}{n}\right)^n \xrightarrow{n\to\infty} e^{-\lambda}$$

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



Compound interest on a loan

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

What is the ratio between b(n,p,k) and b(n,p,k-1) ?

$$\frac{b(n,p,k)}{b(n,p,k-1)} = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{\dfrac{n!}{k!(n-k)!} \quad p}{\dfrac{n!}{(k-1)!(n-k+1)!} \quad (1-p)} = \frac{(n-k+1)p}{k(1-p)} = *$$

$$\text{Plugging in } p = \frac{\lambda}{n} \text{ we get}$$

$$* = \frac{(n-k+1)(\lambda / n)}{k\left(1 - \dfrac{\lambda}{n}\right)} = \frac{(n-k+1)\lambda}{k(n-\lambda)} \xrightarrow{n\to\infty} \frac{\lambda}{k}$$

The poisson distribution

$$b(n,p,k) = b(n,p,0) \cdot \frac{b(n,p,1)}{b(n,p,0)} \cdot \frac{b(n,p,2)}{b(n,p,1)} \cdots \frac{b(n,p,k)}{b(n,p,k-1)} \xrightarrow{n\to\infty} e^{-\lambda} \cdot \frac{\lambda}{1} \cdot \frac{\lambda}{2} \cdots \frac{\lambda}{k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

Consider a cluster with n servers that recieves n requests per second and assigns requests to servers randomly

n=1 one server, recieving exactly 1 request per second

n=2 two servers, each recieved 1 RPS in expectation

n=1000, 1000 servers, each recieves 1RPS in expectation

the probability a server recieves 4 requests is well approximated by the poisson with lambda=1 and k=4

The exponential distribution

How long between two consecutive envents?

$$\Pr(i=1) = p = \frac{\lambda}{n}$$

$$\Pr(i=2) = p(1-p) = \frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)$$

$$\Pr(i=k) = p(1-p)^k = \frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)^k$$

as we have n bins per second, time t corresponds to bin tn

$$\Pr(i=tn) = p(1-p)^{tn} = \frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)^{tn} = \frac{\lambda}{n}\left(\left(1-\frac{\lambda}{n}\right)^{n}\right)^{t}$$

$$\text{if t } \geq 0: f(t) = \lim_{n\to\infty}\frac{\Pr(i=tn)}{1/n} = \lambda\left(\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}\right)^{t} = \lambda e^{-\lambda t}$$

$$F(t) = \int_{-\infty}^{t} f(s)ds = \int_{0}^{t} f(s)ds = 1-e^{-\lambda t}$$

$$\text{if t<0: } \; f(t) = 0, \; F(t) = 0$$

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?

Time length of T corresponds to a rate of lambda=T*100

$$e^{-\lambda}\frac{\lambda^k}{k!}$$

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?

using the cdf for the exponential distribution:

$$\Pr(s \geq t) = 1 - F(t) = 1 - \left(1 - e^{-\lambda t}\right) = e^{-100t}$$

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?