ENGR 1100

Probability

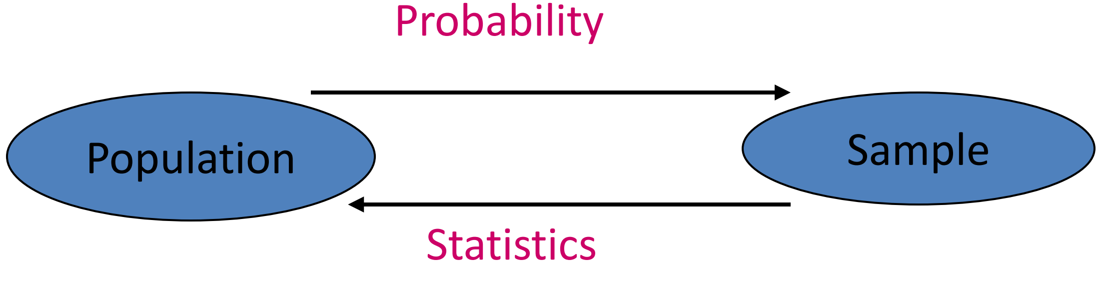
- Random Variables
- Discrete vs. Continuous Distributions

Learning Objectives

- Understand the difference between probability and statistics
- Common discrete and continuous distribution
  - Binomial
  - Poisson
  - Exponential

Why Learn Probability & Statistics?

- Nothing in life is certain. In everything we do, we gauge the chances of successful outcomes, from business to medicine to the weather
- A probability provides a quantitative description of the chances or likelihoods associated with various outcomes
- It provides a bridge between descriptive and inferential statistics



Probabilistic vs Statistical Reasoning

- Suppose I know exactly the proportions of car makes in California. Then I can find the probability that the first car I see in the street is a Ford. This is probabilistic reasoning as I know the population and predict the sample.

- Now suppose that I do not know the proportions of car makes in California, but would like to estimate them. I observe a random sample of cars in the street and then I have an estimate of the proportions of the population. This is statistical reasoning

Experiments and Events

- An **experiment** is the process by which an observation (or measurement) is obtained.
- An **event** is an outcome of an experiment, usually denoted by a capital letter.
- **Example:**
  - **Experiment: Record an age**
    - A: person is 30 years old
    - B: person is older than 65
  - **Experiment: Toss a die**
    - A: observe an odd number
    - B: observe a number greater than 2

The Probability of an Event

- The probability of an event A measures "how often" A will occur. We write **P(A).**
- Suppose that an experiment is performed *n* times. The relative frequency for an event A is

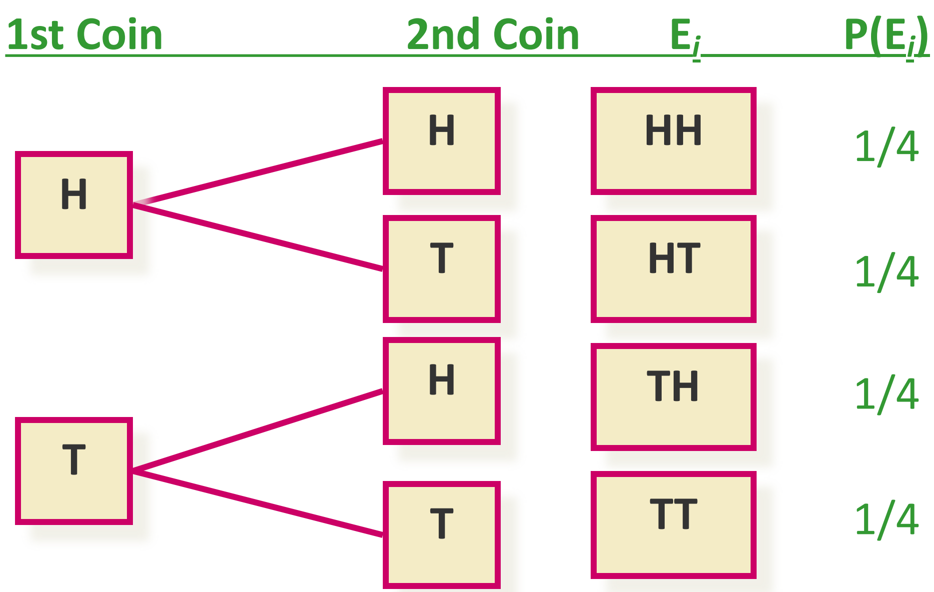
$$\frac{\text{Number of times A occurs}}{n} = \frac{f}{n}$$

- If we let *n* get infinitely large,

$$\boxed{P(A) = \lim_{n \to \infty} \frac{f}{n}}$$

Example 1

Toss a fair coin twice. What is the probability of observing at least one head?



P(at least 1 head)

$= P(E_1) + P(E_2) + P(E_3)$

$= 1/4 + 1/4 + 1/4 = 3/4$

Random Variables

- A quantitative variable *x* is a **random variable** if the value that it assumes, corresponding to the outcome of an experiment is a chance or random event.
- Random variables can be **discrete** or **continuous.**
  - A discrete random variable can take a countable number of values.
    - ✓ Number of people who age 60 and up
  - A continuous random variable can take any value along a given interval of a number line.
    - ✓ The time a tourist stays at the top once s/he gets there

Discrete Random Variables

The **probability distribution for a discrete random variable *x*** resembles the relative frequency distributions. It is a graph, table or formula that gives the possible values of *x* and the probability *p(x)* associated with each value.

We must have

$$0 \leq p(x) \leq 1 \text{ and } \sum p(x) = 1$$

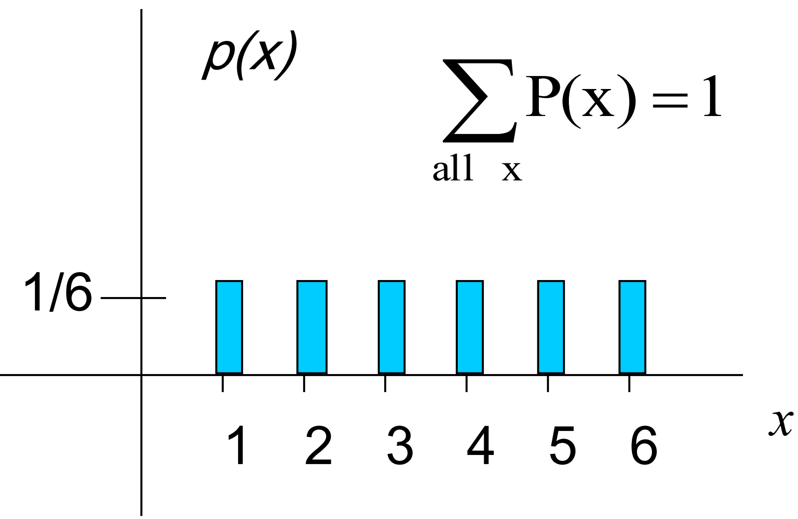
Example 1: roll of a die

Probability Mass Function (pmf)

x	$p(x)$
1	$p(x=1)=1/6$
2	$p(x=2)=1/6$
3	$p(x=3)=1/6$
4	$p(x=4)=1/6$
5	$p(x=5)=1/6$
6	$p(x=6)=1/6$

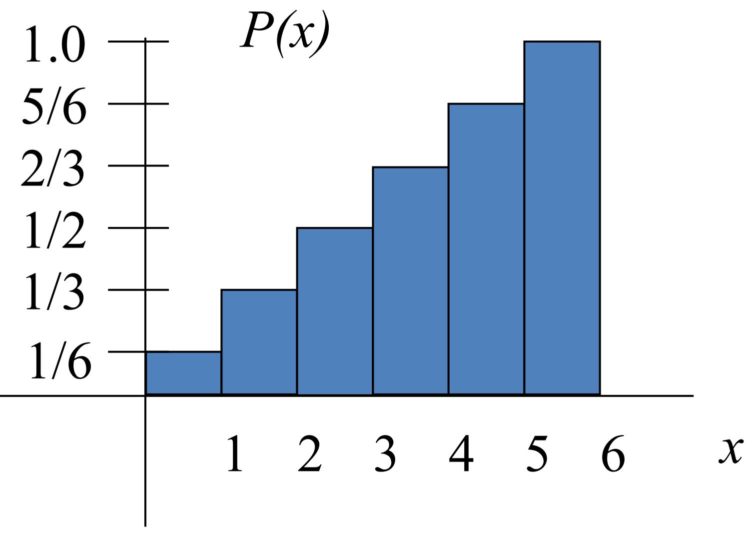
1.0

$$\sum_{\text{all } x} P(x) = 1$$



Cumulative distribution function (CDF)

x	*P(x≤A)*
1	*P(x≤1)*=1/6
2	*P(x≤2)*=2/6
3	*P(x≤3)*=3/6
4	*P(x≤4)*=4/6
5	*P(x≤5)*=5/6
6	*P(x≤6)*=6/6



The Mean and Standard Deviation

Let *x* be a **discrete random** variable with probability distribution *p(x)*. Then the mean, variance and standard deviation of *x* are given as

$$\text{Mean}: \mu = \sum x p(x)$$

$$\text{Variance}: \sigma^2 = \sum (x-\mu)^2 p(x)$$

$$\text{Standard deviation}: \sigma = \sqrt{\sigma^2}$$

Example

Toss a fair coin 3 times and record *x* the number of heads.

x	$p(x)$	$xp(x)$	$(x-\mu)^2p(x)$
0	1/8	0	$(-1.5)^2(1/8)$
1	3/8	3/8	$(-0.5)^2(3/8)$
2	3/8	6/8	$(0.5)^2(3/8)$
3	1/8	3/8	$(1.5)^2(1/8)$

$$\mu = \sum xp(x) = \frac{12}{8} = 1.5$$

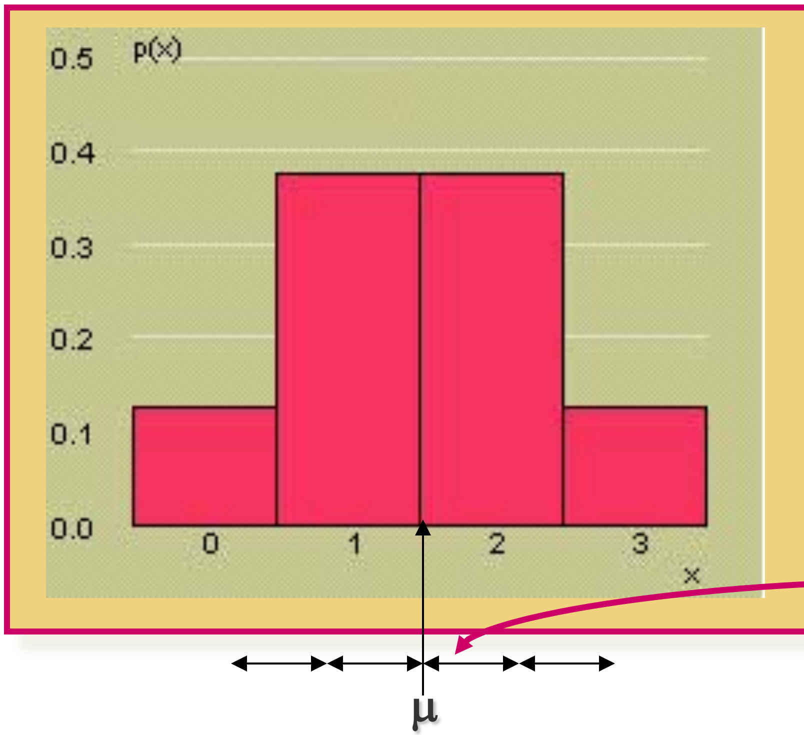
$$\sigma^2 = \sum (x-\mu)^2 p(x)$$

$$\sigma^2 = .28125 + .09375 + .09375 + .28125 = .75$$

$$\sigma = \sqrt{.75} = .688$$

Example

The probability distribution for *x* the number of heads in tossing 3 fair coins.



- Shape? Symmetric; mound-shaped
- Outliers? None
- Center? $\mu = 1.5$
- Spread? $\sigma = .688$

Review Question

Two dice are rolled and the sum of the face values is six? What is the probability that at least one of the dice came up a 3?

a. 1/5
b. 2/3
c. 1/2
d. 5/6
e. 1.0

COMMON DISCRETE PROBABILITY DISTRIBUTIONS

Binomial Distribution

- A fixed number of observations (trials), **n**
  - e.g., 15 tosses of a coin; 20 patients; 1000 people surveyed
- A binary outcome
  - e.g., head or tail in each toss of a coin; disease or no disease
  - Generally called “success” and “failure”
  - Probability of success is p, probability of failure is 1 – p
- Constant probability for each observation, **p**
  - e.g., Probability of getting a tail is the same each time we toss the coin

Binomial Distribution Example

<u>5 coin tosses</u>. What's the probability that you flip exactly 3 heads in 5 coin tosses?

Solution:

n = number of trials

Binomial Coefficients

$$\binom{n}{X} p^X (1-p)^{n-X}$$

X = # successes out of n trials

p = probability of success

$1\text{-}p$ = probability of failure

Binomial distribution: example

If I toss a coin 20 times, what's the probability of getting exactly 10 heads?

$$\binom{20}{10}(.5)^{10}(.5)^{10} = .176$$

Binomial: Mean & Variance

If X follows a binomial distribution with parameters n and p: ***X ~ Bino (n, p)***

Then:

$E(X) = np$

$Var(X) = np(1-p)$

$SD(X) = \sqrt{np(1-p)}$

Note: the variance will always lie between

0*N-.25 *N

p(1-p) reaches maximum at p=.5

P(1-p)=.25

Practice Problem

1. You are performing a cohort study. If the probability of developing disease in the exposed group is .05 for the study duration, then if you (randomly) sample 500 exposed people, how many do you expect to develop the disease? Give a margin of error (+/- 1 standard deviation) for your estimate.

2. What's the probability that **at most** 10 exposed people develop the disease? Write your formula without calculation.

Answer

1. How many do you expect to develop the disease? Give a margin of error (+/- 1 standard deviation) for your estimate.

   X ~ Binomial (500, .05)

   E(X) = 500 (.05) = 25

   Var(X) = 500 (.05) (.95) = 23.75

   StdDev(X) = square root (23.75) = 4.87

   So, on average 25 $\pm$ 4.87

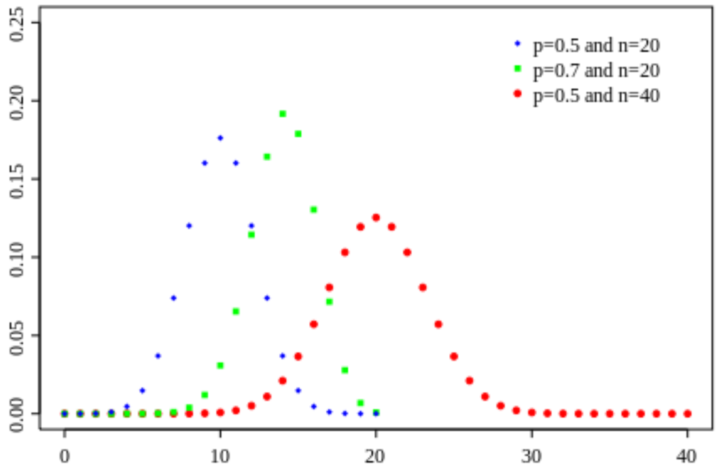
2. Compute the probability that <u>at most</u> 10 exposed people develop the disease.

   The probability is given by P(X<=10) where X is a binomial with parameter n=500 and p =0.05

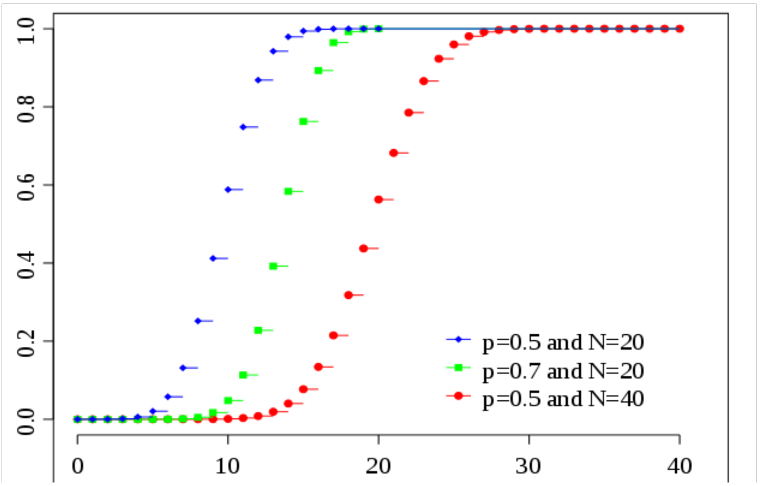
MATLAB: Binomial

- The binomial distribution is used for processes that have a success or fail probability and is useful for determining the total probable number of successes

PDF



CDF



- The MATLAB call for the pdf and cdf for the binomial distributions are

  >> binocdf(x,N,p)

  >> binopdf(x,N,p)

MATLAB Practice

Use MATLAB to visualize a Binomial distribution with parameter N=20, p=0.4.

Poisson Distribution

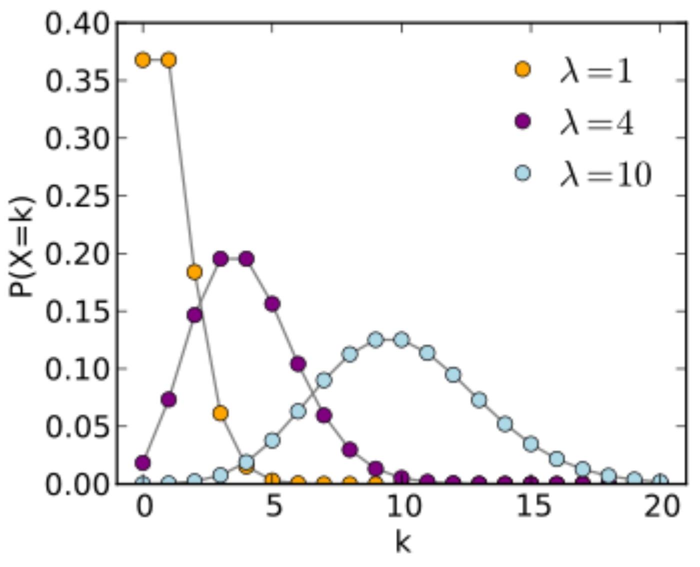
The Poisson distribution is used to model the number of events occurring within a given time interval. The formula for the Poisson probability density (mass) function is

$$\boldsymbol{p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}}$$

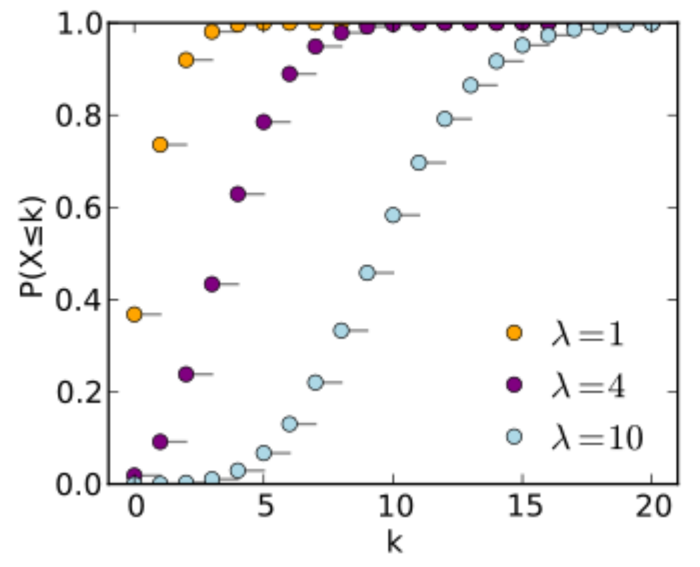
λ is the shape parameter which indicates the average number of events in the given time interval.

Common Distributions: Poisson

PDF



CDF



- The MATLAB for Poisson distributions are

```
>> poisscdf(x,lambda)
>> poisspdf(x,lambda)
>> poissinv(P,lambda)
>> poissrnd(lambda, M, N)
```

Poisson Distributions in MATLAB

- Compute probability density function

```
>> poisspdf(x,lambda)
```

- Compute cumulative probability function

```
>> poisscdf(x,lambda)
```

- Compute the inverse of Poisson

```
>> poissinv(P,lambda)
```

Example: If the average number of defects (λ) is 2, what is the 95th percentile of the number of defects?

```
poissinv(0.95,2)
ans = 5
```

- Generate Poisson random variables

```
>> poissrnd(lambda, M,N)
```

Example - Poisson Distribution

Arrivals at a bus-stop follow a Poisson distribution with an average of 4.5 every quarter of an hour.

Calculate the probability of fewer than 3 arrivals in a quarter of an hour.

Example - Poisson Distribution

The probabilities of 0 up to 2 arrivals can be calculated directly from the formula

$$p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$$p(0) = \frac{e^{-4.5}4.5^{0}}{0!} = 0.0111$$

$$p(1) = \frac{e^{-4.5}4.5^{1}}{1!} = 0.0499 \qquad p(2) = \frac{e^{-4.5}4.5^{2}}{2!} = 0.1124$$

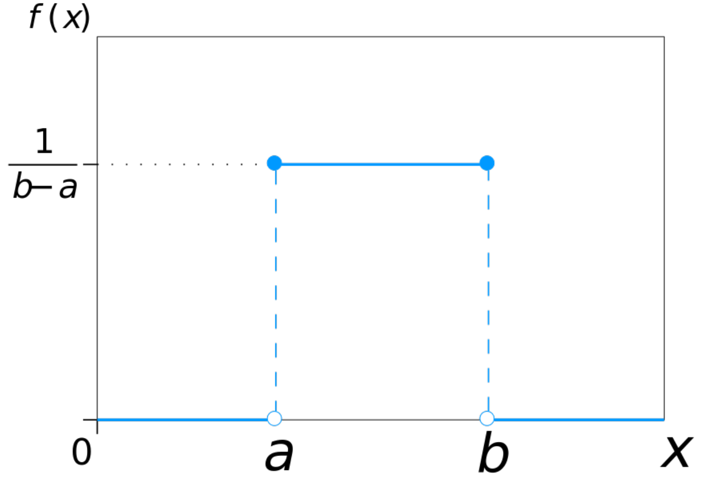
So the probability of fewer than 3 arrivals is 0.1735

COMMON CONTINUOUS PROBABILITY DISTRIBUTIONS

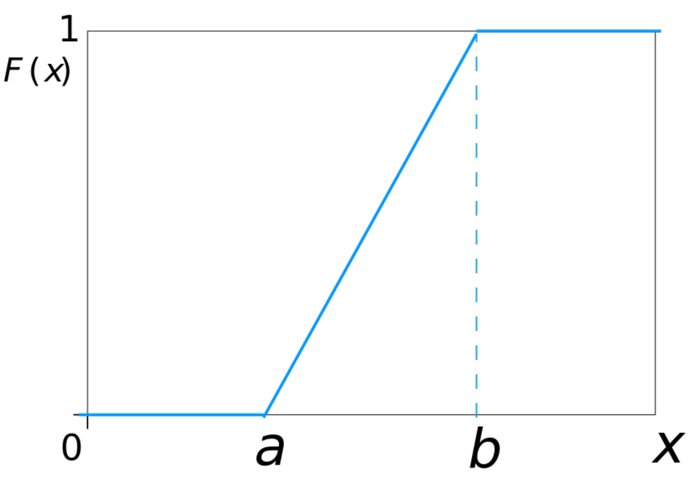
Common Distributions: Uniform

- The most basic distribution is the uniform distribution which sets all probabilities of possible values equal to each other

PDF



CDF



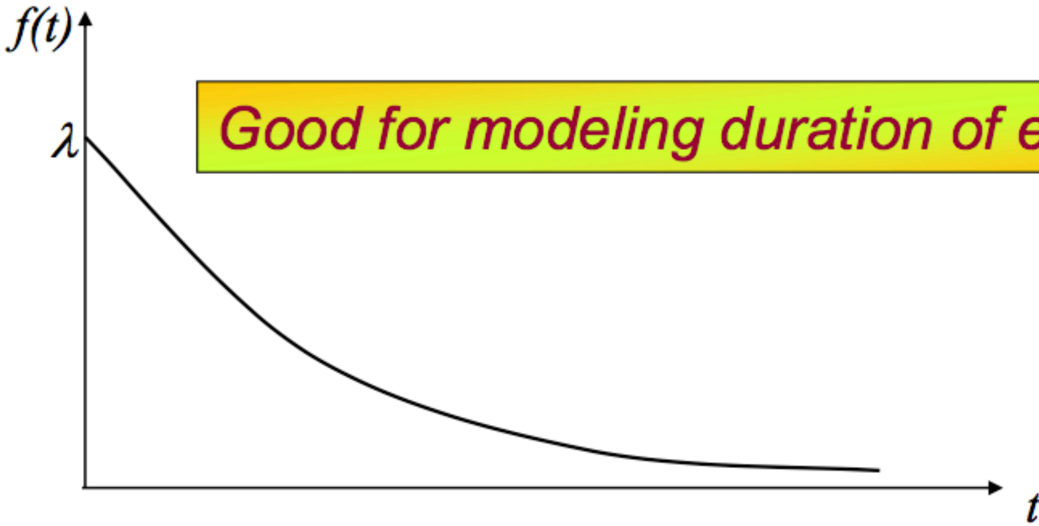
- Uniform variables can either be discrete or continuous
- In MATLAB the command for calling the pdf and cdf of a uniform distribution are unidpdf(), unidcdf(), unifpdf(), and unifcdf()

```
>> unidcdf(x,N)
>> unifpdf(x,a,b)
```

The Exponential Distribution

A random variable T follows the exponential distribution with rate $\lambda > 0$ if its p.d.f is :

$$f_T(t) = \lambda\, e^{-\lambda t} \qquad \text{for} \quad t > 0$$



Good for modeling duration of events

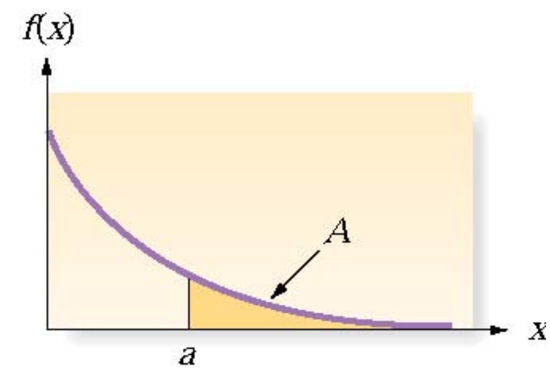
The Exponential Distribution Example

- The life time of an electronic device
- The time between arrivals of two successive buses
- Time until next earthquake occurs

Example: Suppose the waiting time to see the nurse at the student health center is distributed exponentially with a mean of 45 minutes. What is the probability that a student will wait more than an hour to get his or her generic pill?

$$P(W \geq x) = e^{-x/\theta}$$

$$P(W \geq 60) = e^{-60/45} = e^{-1.33} = 0.2645$$



Properties of Exponential Distribution

1. Expected value (mean)

$$E[T] = \int_0^{\infty} tf(t)dt = \int_0^{\infty} t\lambda e^{-\lambda t} dt = 1/\lambda$$

λ = average number of events in unit of time.
$1/\lambda$ = average time until an event occurs.

2. Variance

$$Var[T] = E[T^2] - (E[T])^2 = 1/\lambda^2$$

Properties of Exponential Distribution

- Memory less property

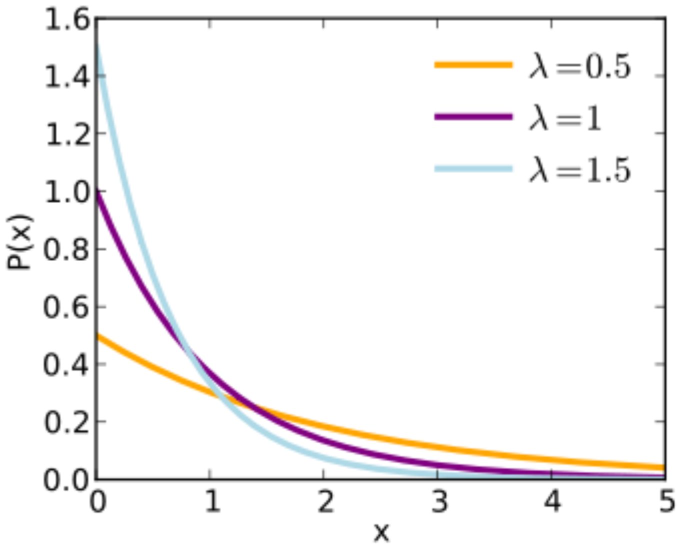
T is exponentially distributed with rate λ.

$$\Pr\{T > t + h \mid T > t\} = \Pr\{T > h\}$$

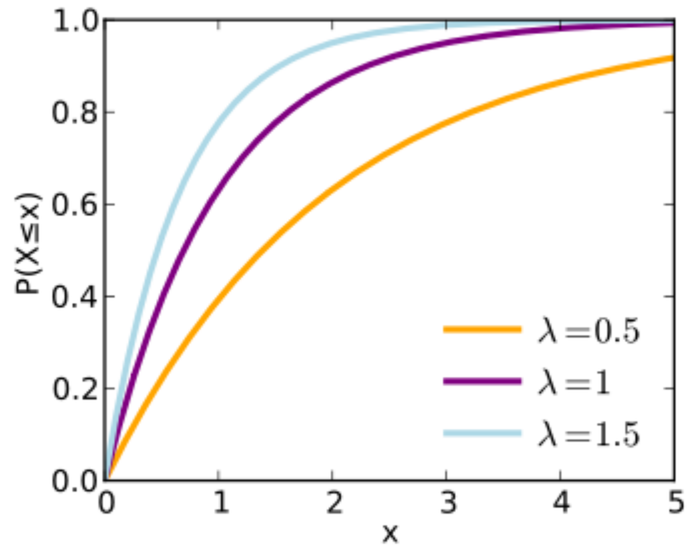
Example: if the battery worked for 7 days. What is the probability that it will work 3 additional days?

Common Distributions: Exponential

PDF



CDF



- The MATLAB function for Exponential distribution are

```
>> expcdf(x,lambda)
>> exppdf(x,lambda)
>> expinv(P,lambda)
>> exprnd(lambda, M, N)
```

Exponential Distributions in MATLAB

- Compute probability density function

```
>> exppdf(x,lambda)
```

- Compute cumulative probability function

```
>> expcdf(x,lambda)
```

- Compute the inverse of Poisson

```
>> expinv(P,lambda)
```

Example: Let the lifetime of light bulbs be exponentially distributed with λ = 700 hours. What is the median lifetime of a bulb?

```
expinv(0.5,700)
 ans = 5
```

- Generate Poisson random variables

```
>> exprnd(lambda, M,N)
```

Example- Exponential Distribution

The time required to repair a machine is an exponential random variable with rate $\lambda = 0.5$.

1. What is the probability that a repair time exceeds 2 hours
2. What is the probability that the repair time will take at least 4 hours given that the repair man has been working on the machine for 3 hours?

Let T : repair time $\Rightarrow f(t) = 0.5\ e^{-0.5t}$ for $t > 0$

1. $P\{T \geq 2\} = e{-}0.5(2) = e{-}1 = 0.36788$
2. $P\{T \geq 4 \mid T \geq 3\} = P\{T \geq 1\} = e^{-0.5(1)} = 0.60653$