Regents Park Publishers

Data Analytics

T1LM 5

v.2

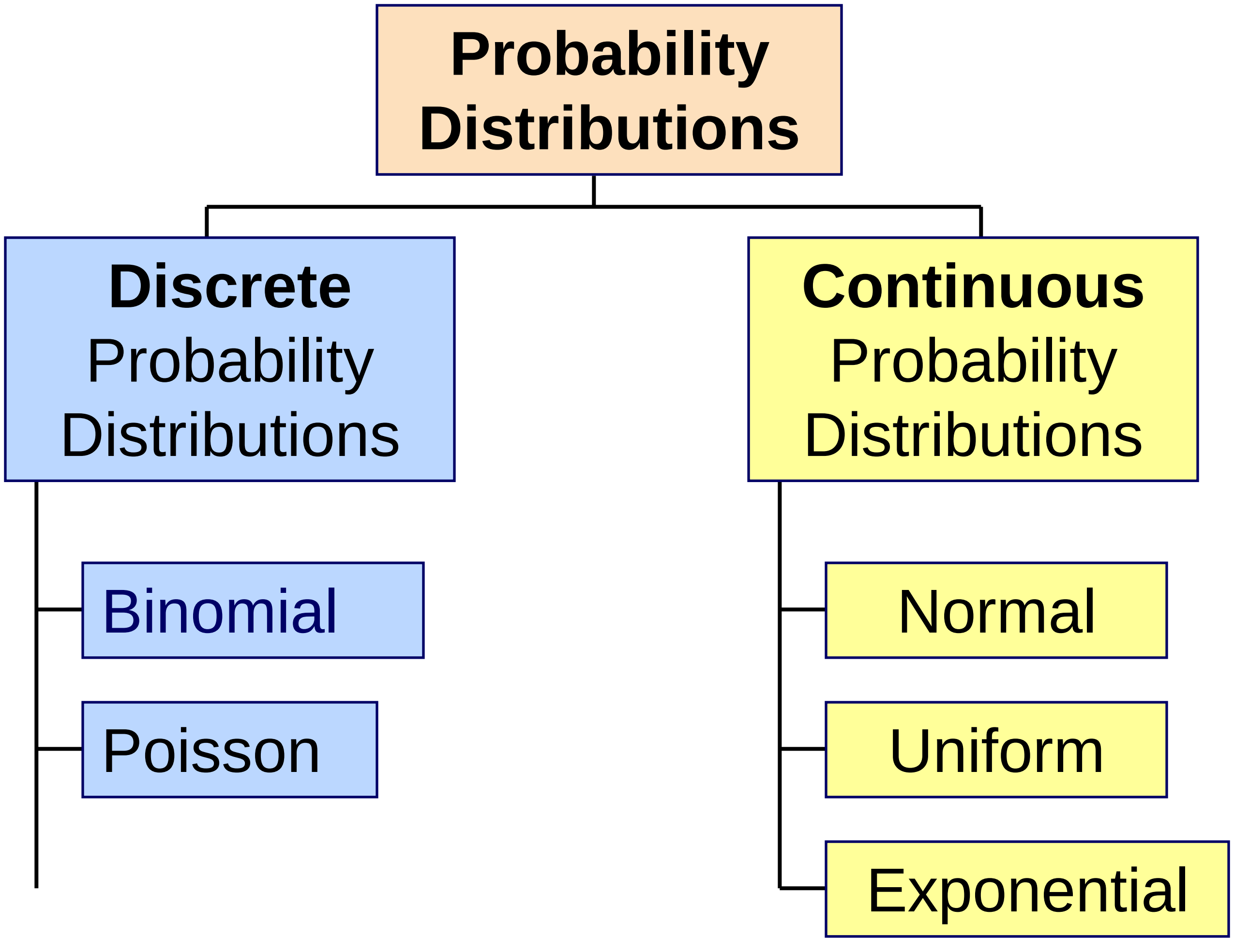
Probability Distribution

Chapter Goals

After completing this chapter, you should be able to:

- Apply the binomial distribution to applied problems
- Compute probabilities for the Poisson distribution
- Find probabilities using a normal distribution table and apply the normal distribution to business problems
- Recognize when to apply the uniform and exponential distributions

Probability Distributions



Discrete Probability Distributions

- A **discrete random variable** is a variable that can assume only a countable number of values

  Many possible outcomes:

  - number of complaints per day
  - number of TVs in a household
  - number of rings before the phone is answered

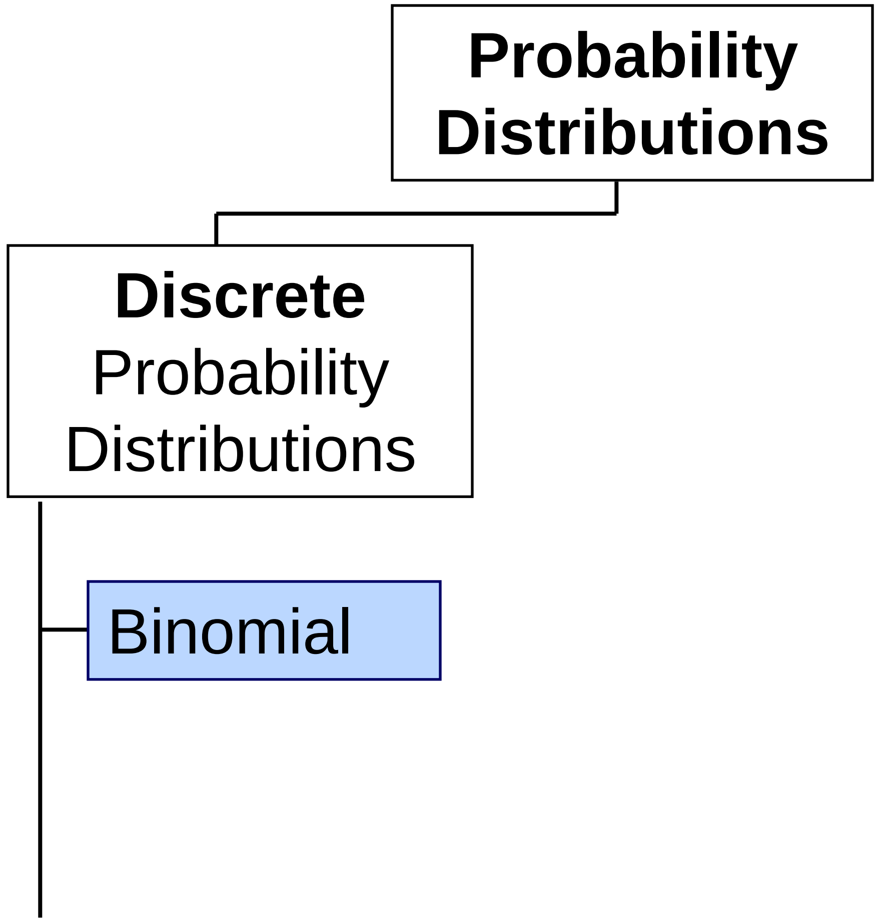
  Only two possible outcomes:

  - defective: yes or no
  - spreads peanut butter first vs. spreads jelly first

Continuous Probability Distributions

- A continuous random variable is a variable that can assume any value on a continuum (can assume an uncountable number of values)
  - thickness of an item
  - time required to complete a task
  - temperature of a solution
  - height, in inches
- These can potentially take on any value, depending only on the ability to measure accurately.

The Binomial Distribution



The Binomial Distribution

- Characteristics of the Binomial Distribution:
  - A trial has only two possible outcomes – “success” or “failure”
  - There is a fixed number, n, of identical trials
  - The trials of the experiment are independent of each other
  - The probability of a success, p, remains constant from trial to trial
  - If p represents the probability of a success, then $(1-p) = q$ is the probability of a failure

Binomial Distribution Settings

- A manufacturing plant labels items as either defective or acceptable
- A firm bidding for a contract will either get the contract or not
- A marketing research firm receives survey responses of “yes I will buy” or “no I will not”
- New job applicants either accept the offer or reject it

Binomial Characteristics

Examples

$$\mu = np = (5)(.1) = 0.5$$

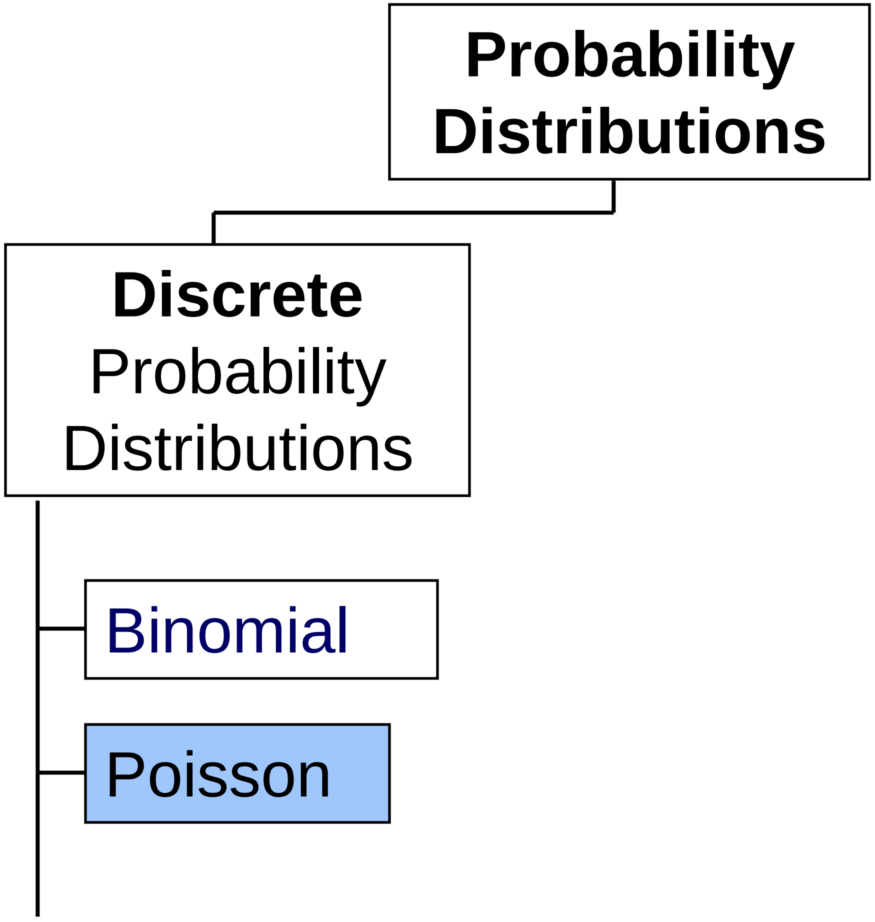
$$\sigma = \sqrt{npq} = \sqrt{(5)(.1)(1-.1)}$$
$$= 0.6708$$

$$\mu = np = (5)(.5) = 2.5$$

$$\sigma = \sqrt{npq} = \sqrt{(5)(.5)(1-.5)}$$
$$= 1.118$$

Path in EXCEL: Formulas to More Functions to Statistical to **BINOM. DIST**

The Poisson Distribution



The Poisson Distribution

- Characteristics of the Poisson Distribution:
  - The outcomes of interest are rare relative to the possible outcomes
  - The average number of outcomes of interest per time or space interval is λ
  - The number of outcomes of interest are random, and the occurrence of one outcome does not influence the chances of another outcome of interest
  - The probability of that an outcome of interest occurs in a given segment is the same for all segments

Poisson Distribution Formula

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

where:

- t = size of the segment of interest
- x = number of successes in segment of interest
- λ = expected number of successes in a segment of unit size
- e = base of the natural logarithm system (2.71828...)

Path in EXCEL: Formulas to More Functions to Statistical to POISSON.DIST

Poisson Distribution Characteristics

- Mean

$$\mu = \lambda t$$

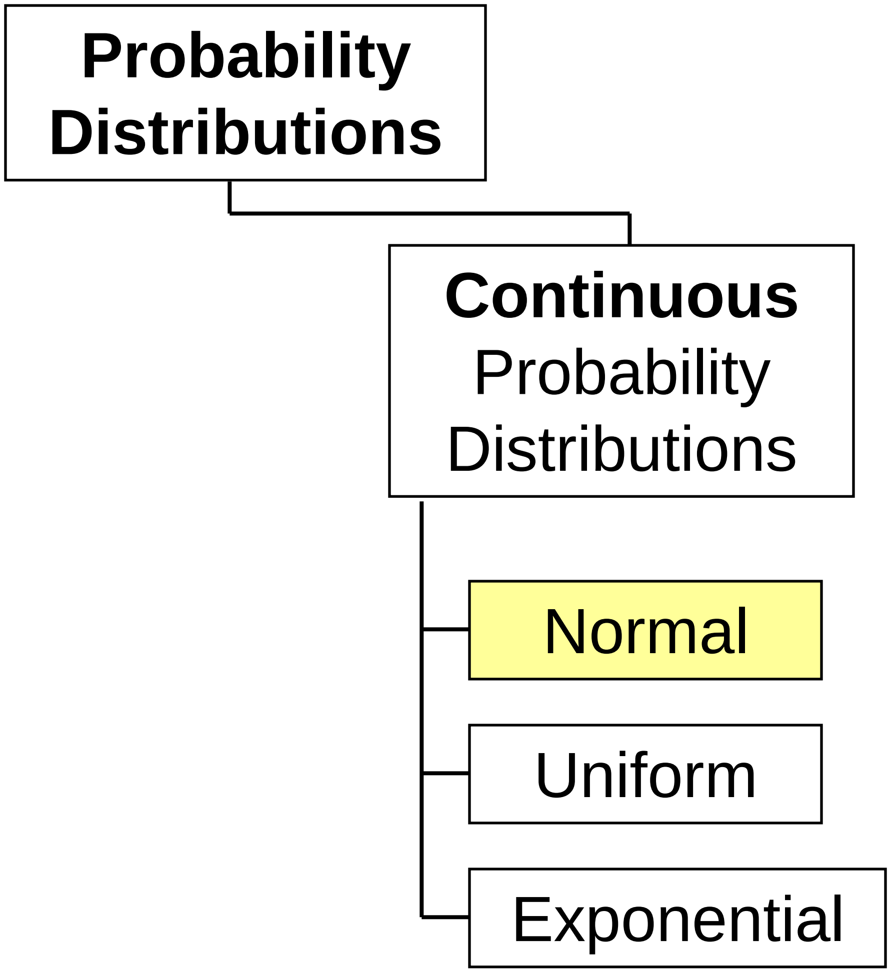
- Variance and Standard Deviation

$$\sigma^2 = \lambda t \qquad \sigma = \sqrt{\lambda t}$$

where λ = number of successes in a segment of unit size
t = the size of the segment of interest

EXCEL: Formulas to More Functions to Statistical to **POISSON. DIST**

The Normal Distribution



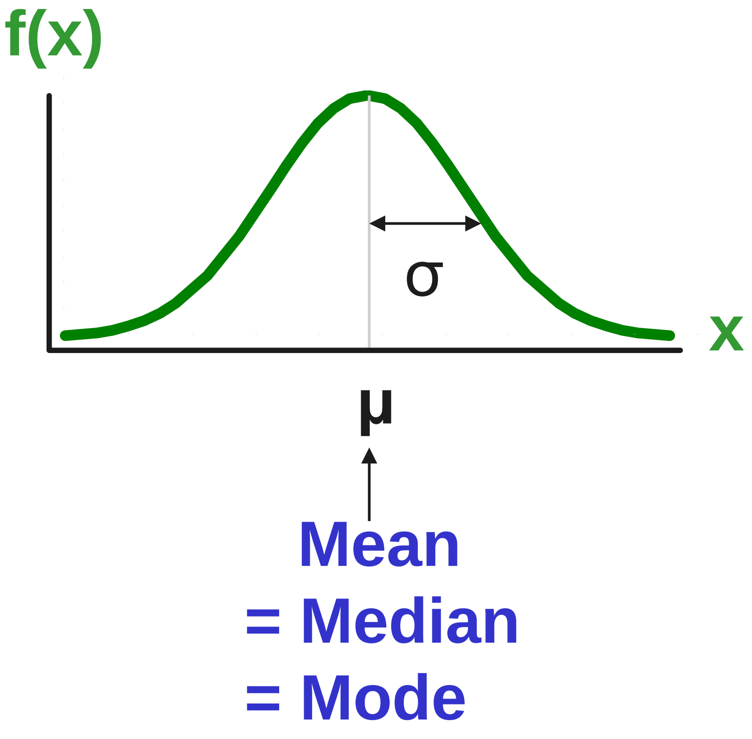
The Normal Distribution

- Bell Shaped
- Symmetrical
- Mean, Median and Mode are Equal

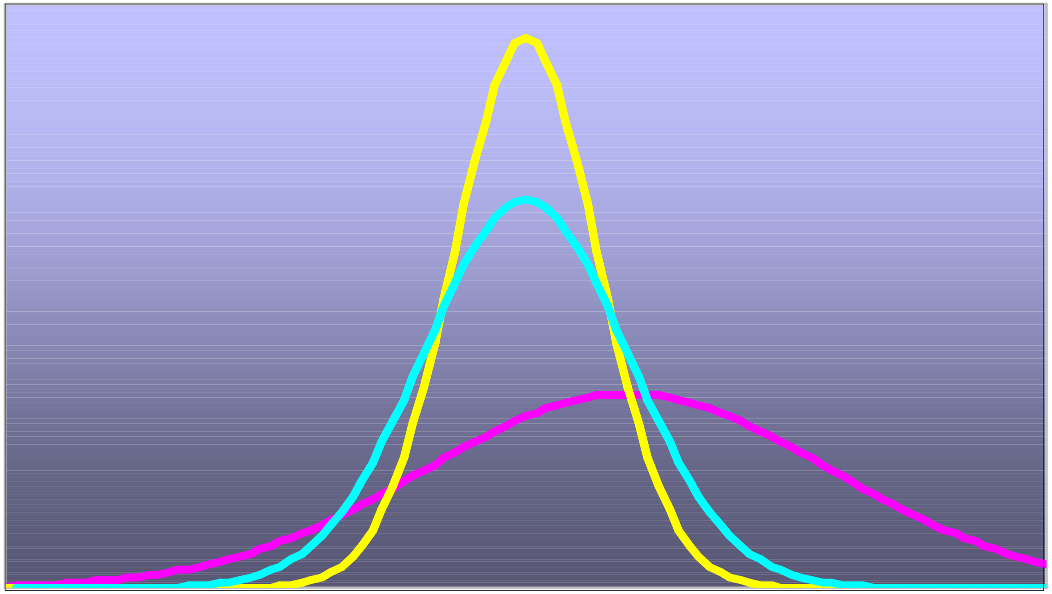
Location is determined by the mean, μ

Spread is determined by the standard deviation, σ

The random variable has an infinite theoretical range: $+\infty$ to $-\infty$

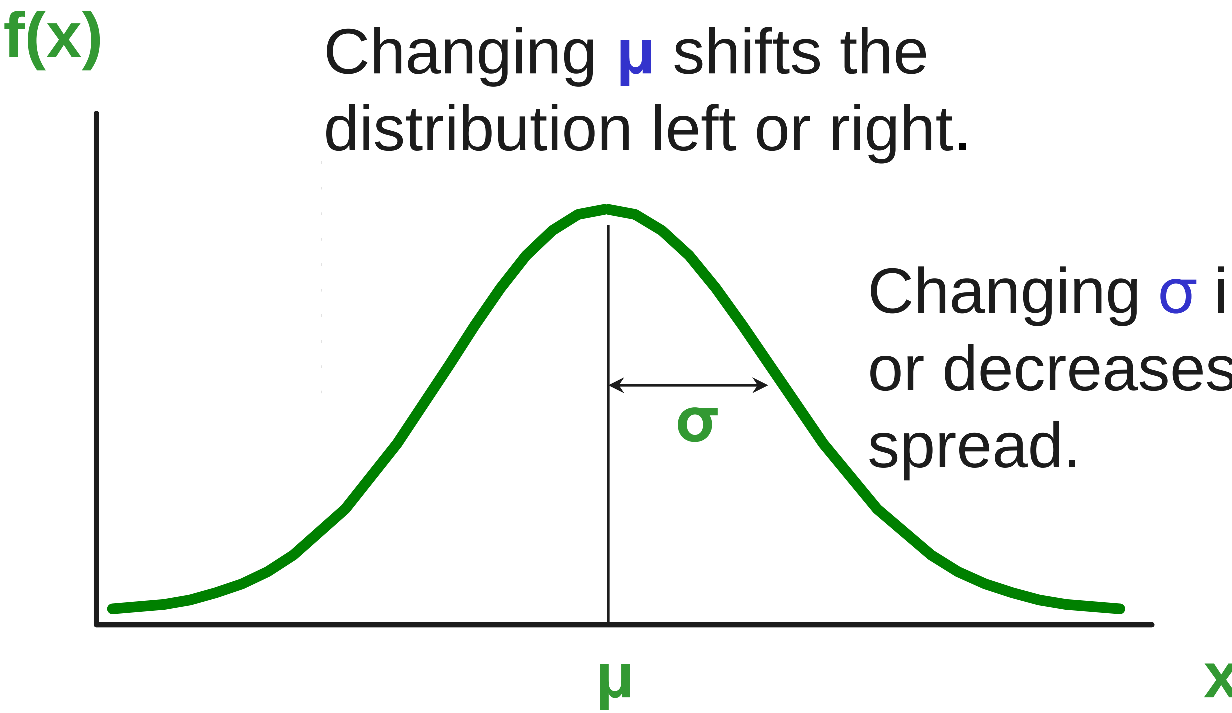


Many Normal Distributions



By varying the parameters μ and σ, we obtain different normal distributions

The Normal Distribution Shape

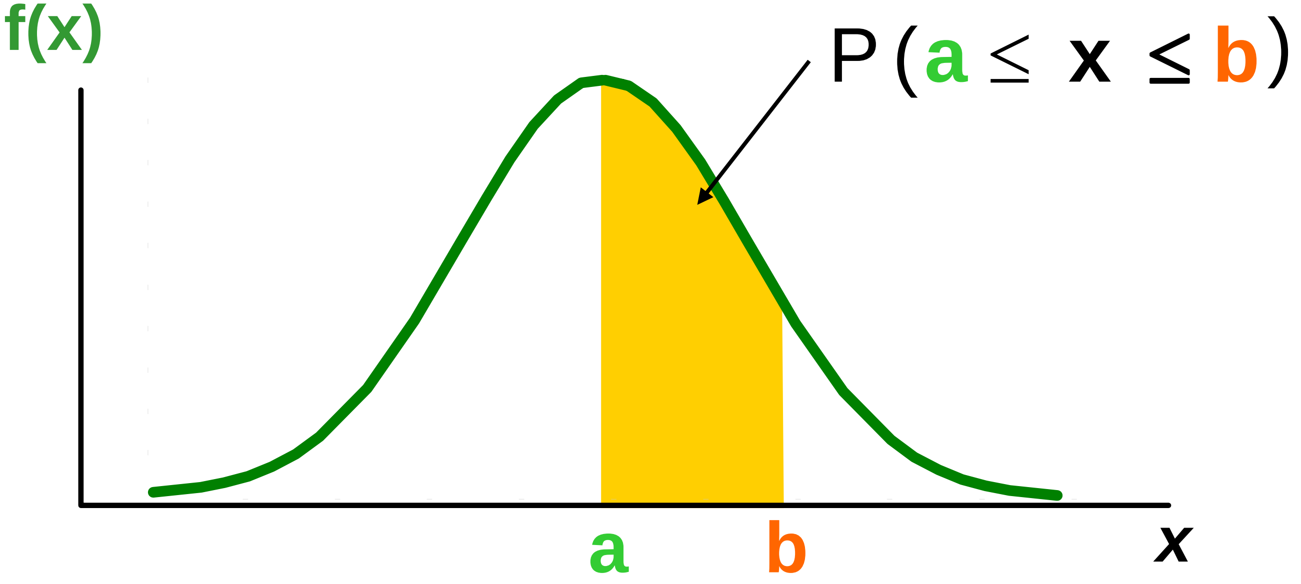


Changing μ shifts the distribution left or right.

Changing σ increases or decreases the spread.

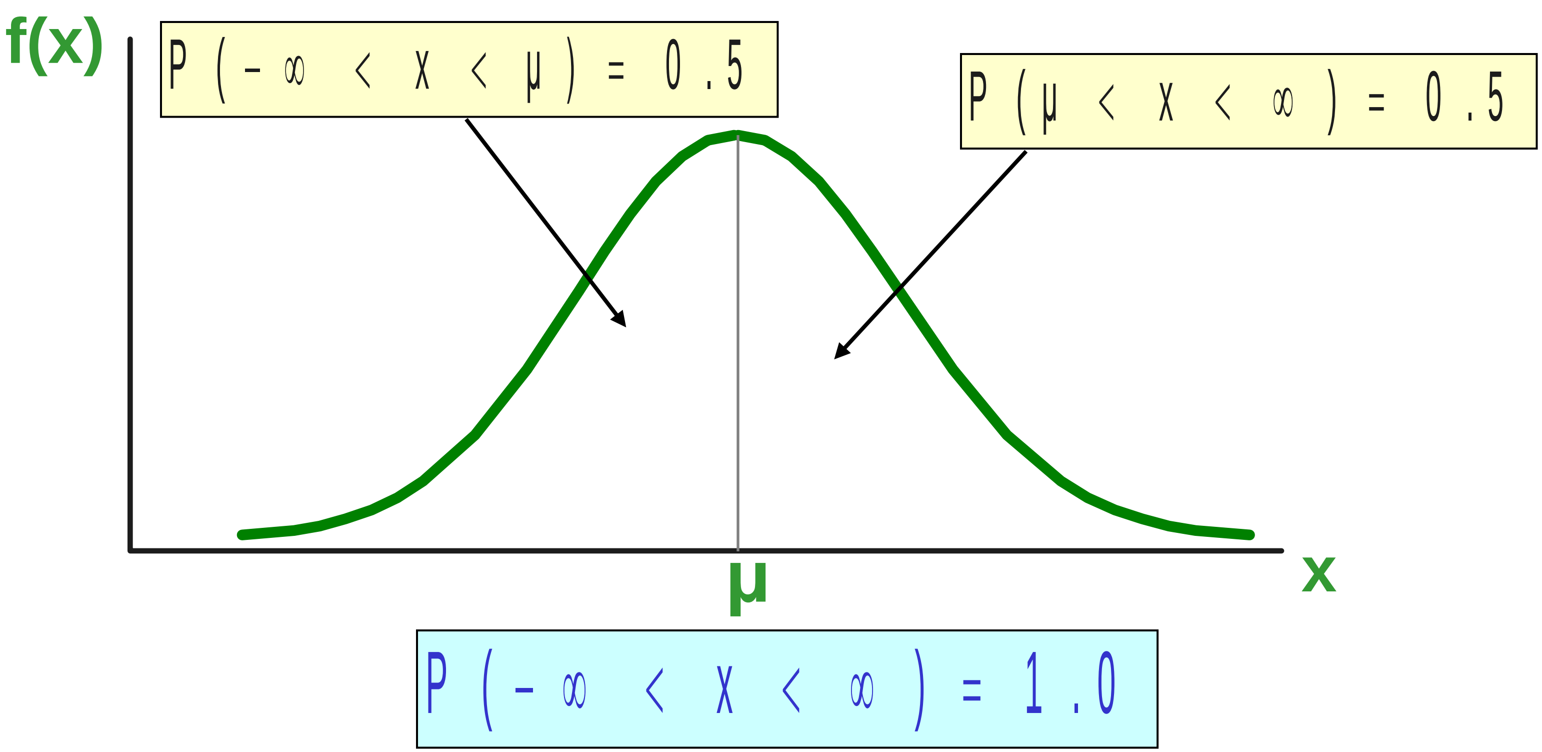
Finding Normal Probabilities

Probability is measured by the area under the curve



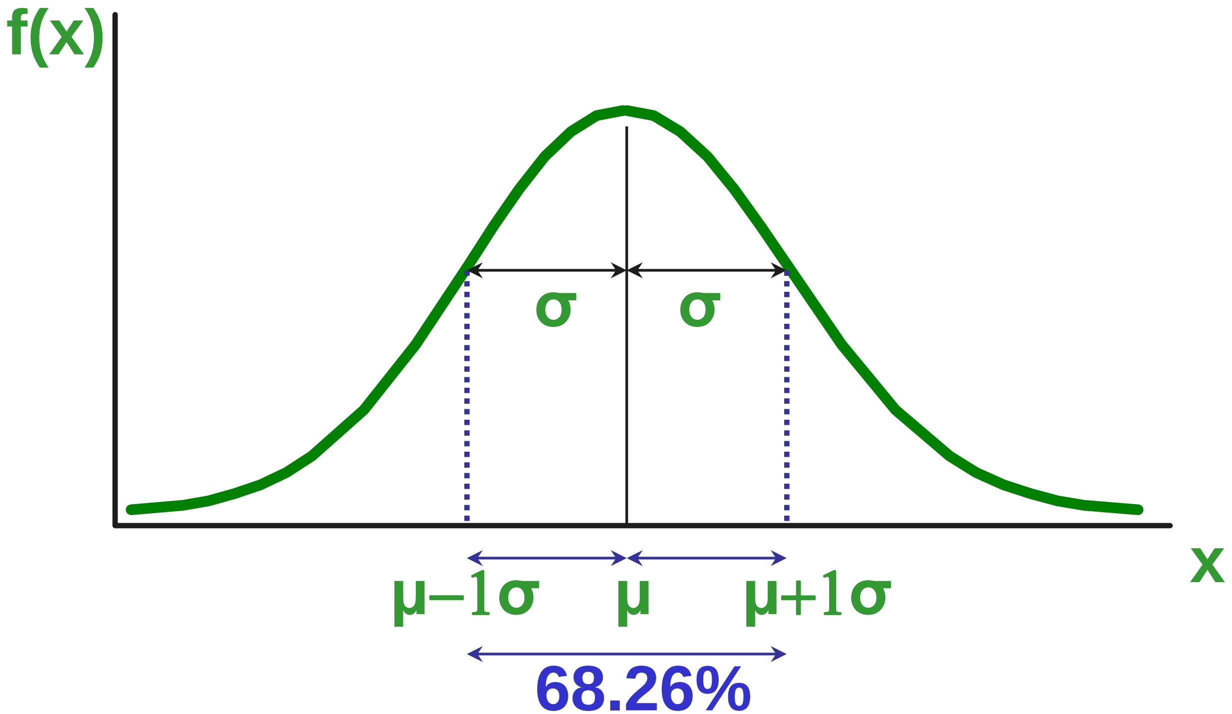
Probability as Area Under the Curve

The total area under the curve is 1.0, and the curve is symmetric, so half is above the mean, half is below



Empirical Rules

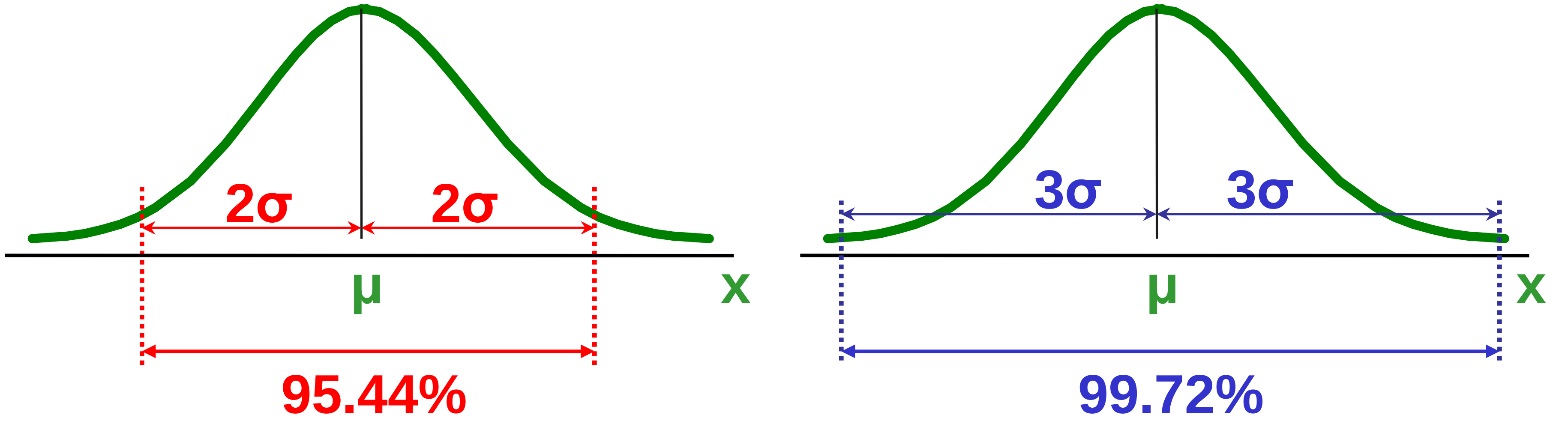
What can we say about the distribution of values around the mean? There are some general rules:



The Empirical Rule

(continued)

- $\mu \pm 2\sigma$ covers about **95.44%** of x's
- $\mu \pm 3\sigma$ covers about **99.72%** of x's



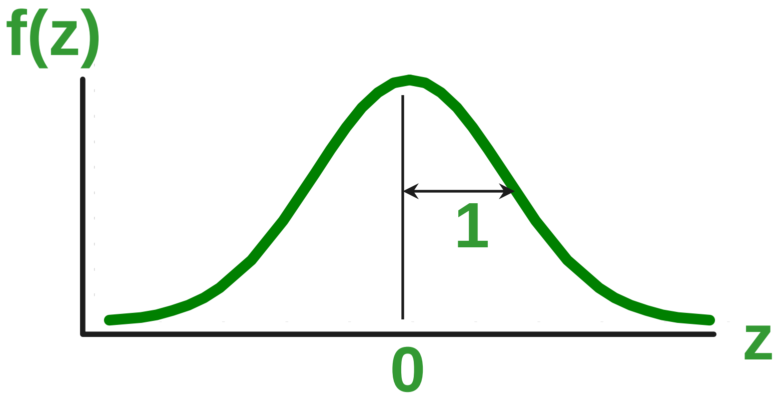
Importance of the Rule

- If a value is about 2 or more standard deviations away from the mean in a normal distribution, then it is far from the mean

- The chance that a value that far or farther away from the mean is highly unlikely, given that particular mean and standard deviation

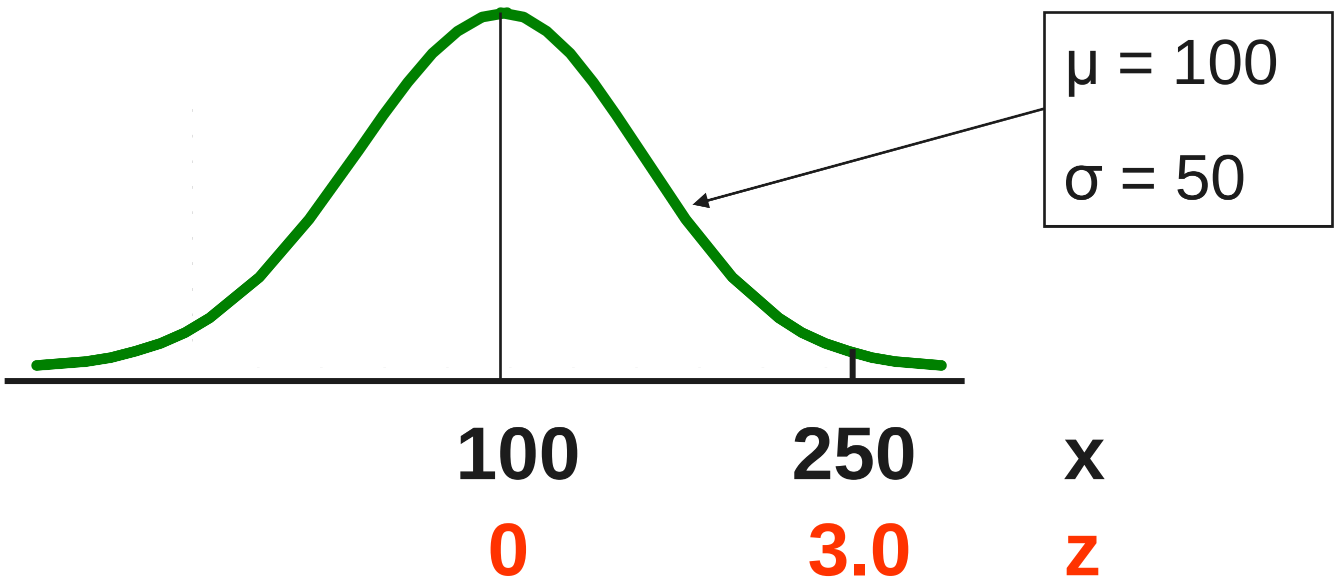
The Standard Normal Distribution

- Also known as the “z” distribution
- Mean is defined to be 0
- Standard Deviation is 1



Values above the mean have positive z-values, values below the mean have negative z-values

Comparing x and z units



Note that the distribution is the same, only the scale has changed. We can express the problem in original units (x) or in standardized units (z)

The Standard Normal

- Any normal distribution (with any mean and standard deviation combination) can be transformed into the standard normal distribution (z)

- Need to transform x units into z units

EXCEL: Formulas to More Functions to Statistical to **NORMS.DIST**

Translation to the Standard Normal Distribution

- Translate from x to the standard normal (the "z" distribution) by subtracting the mean of x and dividing by its standard deviation:

$$\boxed{z = \frac{x - \mu}{\sigma}}$$

Example

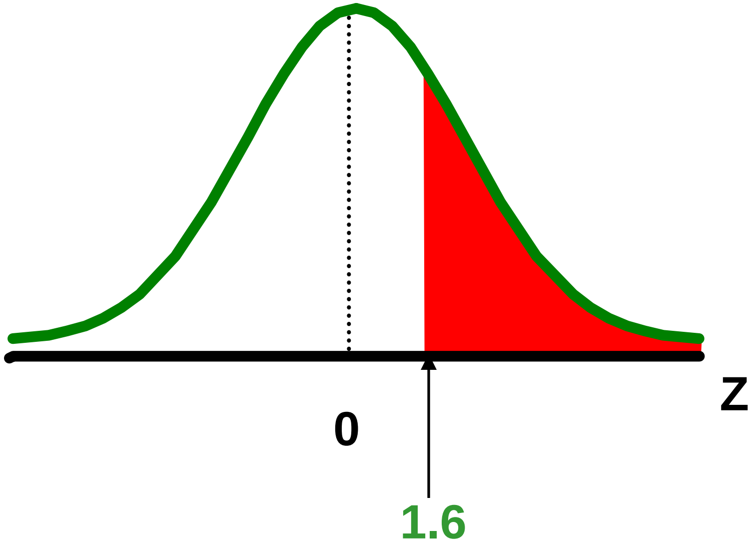
- If x is distributed normally with mean of 100 and standard deviation of 50, the z value for x = 250 is

$$z = \frac{x - \mu}{\sigma} = \frac{250 - 100}{50} = 3.0$$

- This says that x = 250 is three standard deviations (3 increments of 50 units) above the mean of 100.

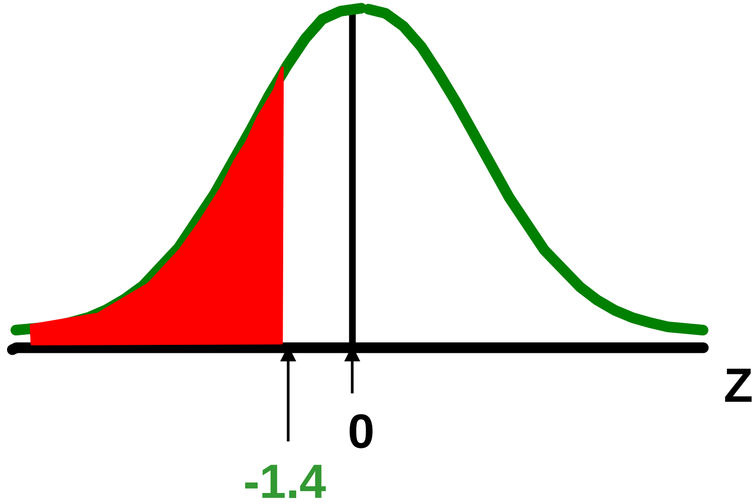
Upper Tail Probabilities

Find P(x > 1.6)



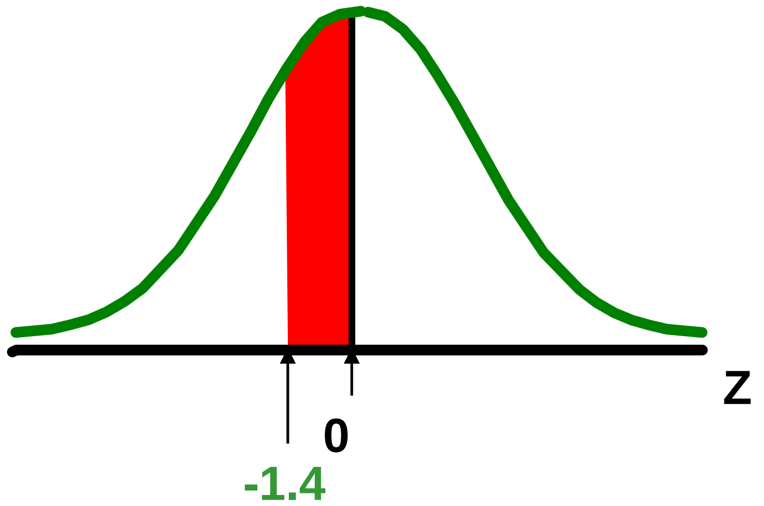
Lower Tail Probabilities

Find P(x < -1.4)

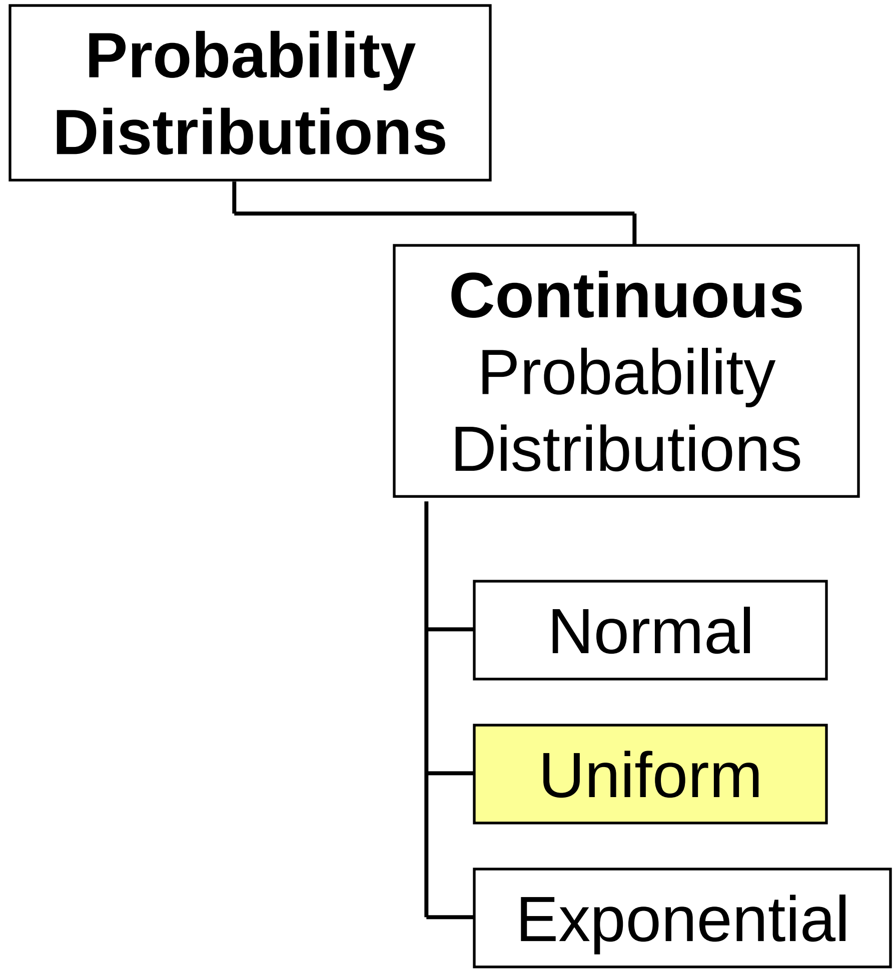


Lower Tail Range of Probabilities

Find P(-1.4 < x < 0)



The Uniform Distribution



The Uniform Distribution

- The uniform distribution is a probability distribution that has equal probabilities for all possible outcomes of the random variable

The Uniform Distribution

(continued)

The Continuous Uniform Distribution:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

where

f(x) = value of the density function at any x value

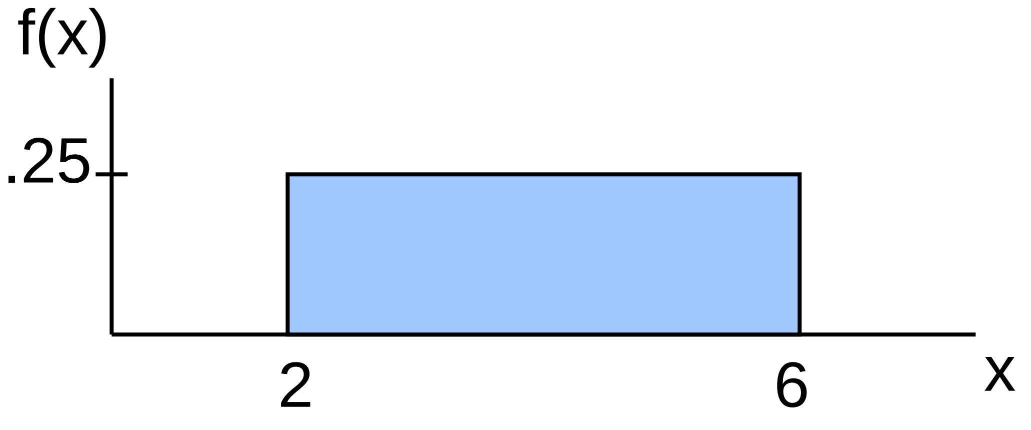
a = lower limit of the interval

b = upper limit of the interval

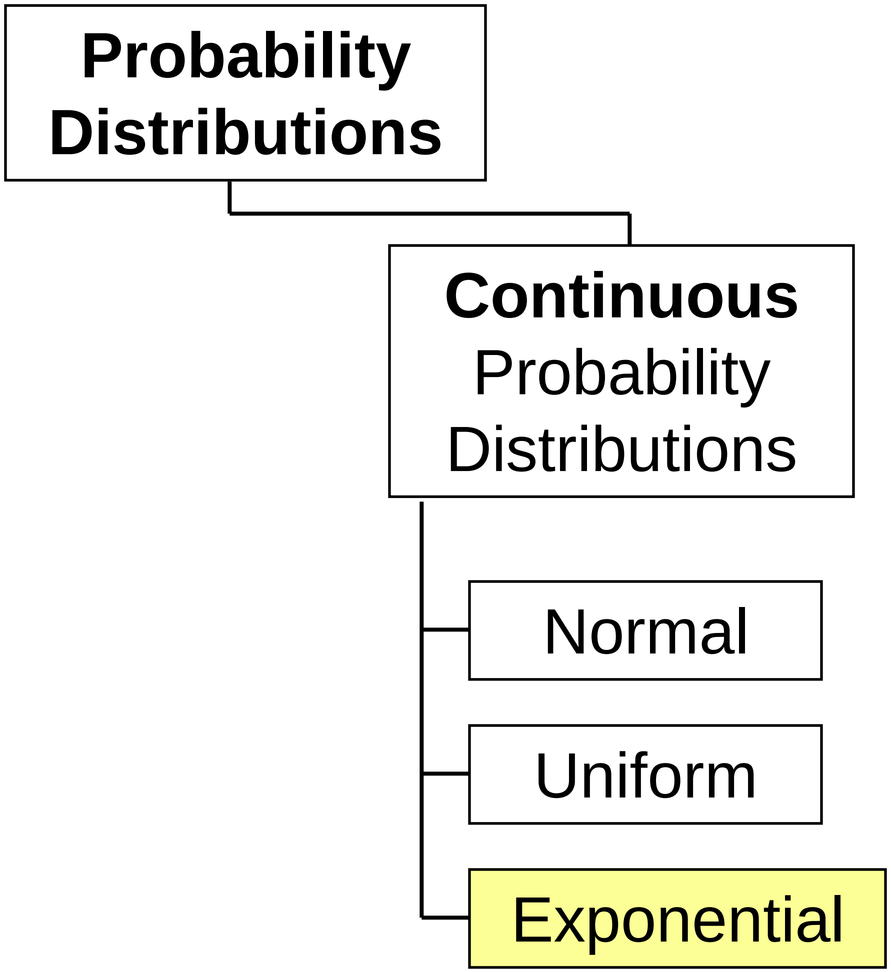
Uniform Distribution

Example: Uniform Probability Distribution Over the range $2 \le x \le 6$:

$$f(x) = \frac{1}{6 - 2} = .25 \quad \text{for} \quad 2 \le x \le 6$$



The Exponential Distribution



The Exponential Distribution

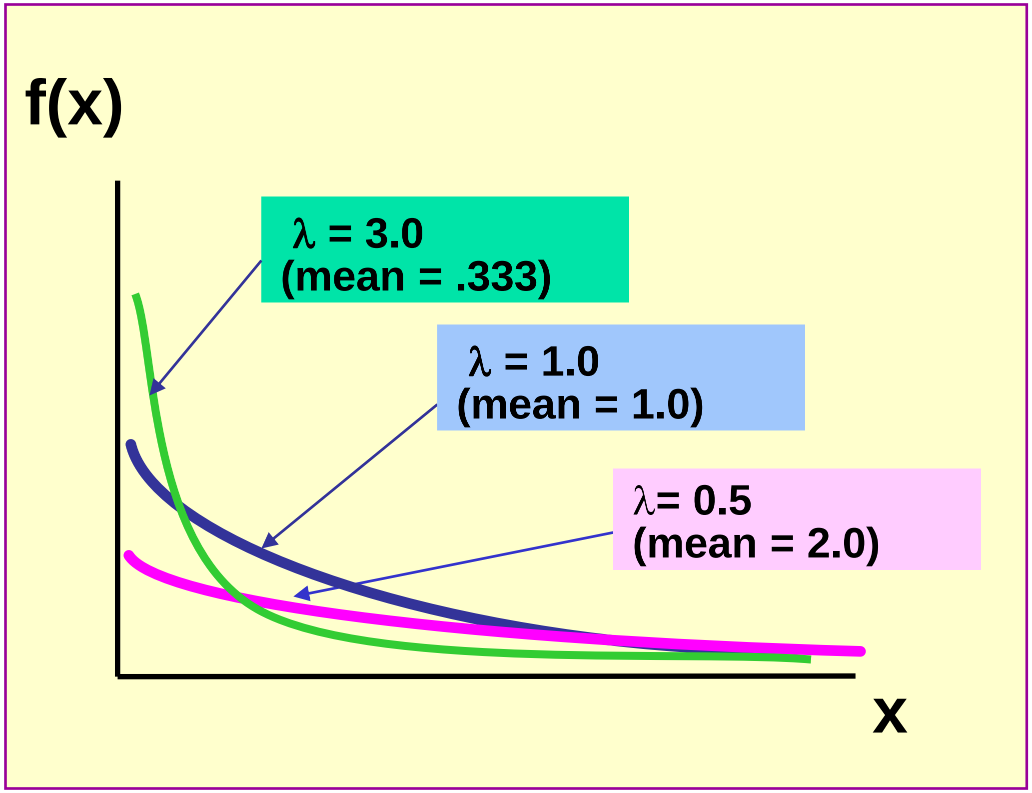
- Used to measure the time that elapses between two occurrences of an event (the time between arrivals)

  - Examples:
    - Time between trucks arriving at an unloading dock
    - Time between transactions at an ATM Machine
    - Time between phone calls to the main operator

Exponential Distribution

(continued)

- Shape of the exponential distribution



Example

Example: Customers arrive at the claims counter at the rate of 15 per hour (Poisson distributed).

What is the probability that the arrival time between consecutive customers is less than five minutes?

- Time between arrivals is exponentially distributed with mean time between arrivals of 4 minutes

  (15 arrivals per 60 minutes, on average)

- $1/\lambda = 4.0$, so $\lambda = .25$
- $P(x < 5) = 1 - e^{-\lambda a} = 1 – e^{-(.25)(5)} =$ **.7135**

EXCEL: Formulas to More Functions to Statistical to **EXP. DIST**

Chapter Summary

- Reviewed key discrete distributions
  - Binomial and Poisson
- Reviewed key continuous distributions
  - Normal, Uniform, Exponential
- Probabilities using formulas and Excel
- Recognized when to apply different distributions

Regents Park Publishers

Data Analytics

T1LM 5

End