

Statistical Methods and Interpretation



Chapter 5

Continuous Random Variables and Probability Distributions



Chapter Goals

After completing this chapter, you should be able to:

- Explain the difference between a discrete and a continuous random variable
- Describe the characteristics of the uniform and normal distributions
- Translate normal distribution problems into standardized normal distribution problems
- Find probabilities using a normal distribution table



Chapter Goals

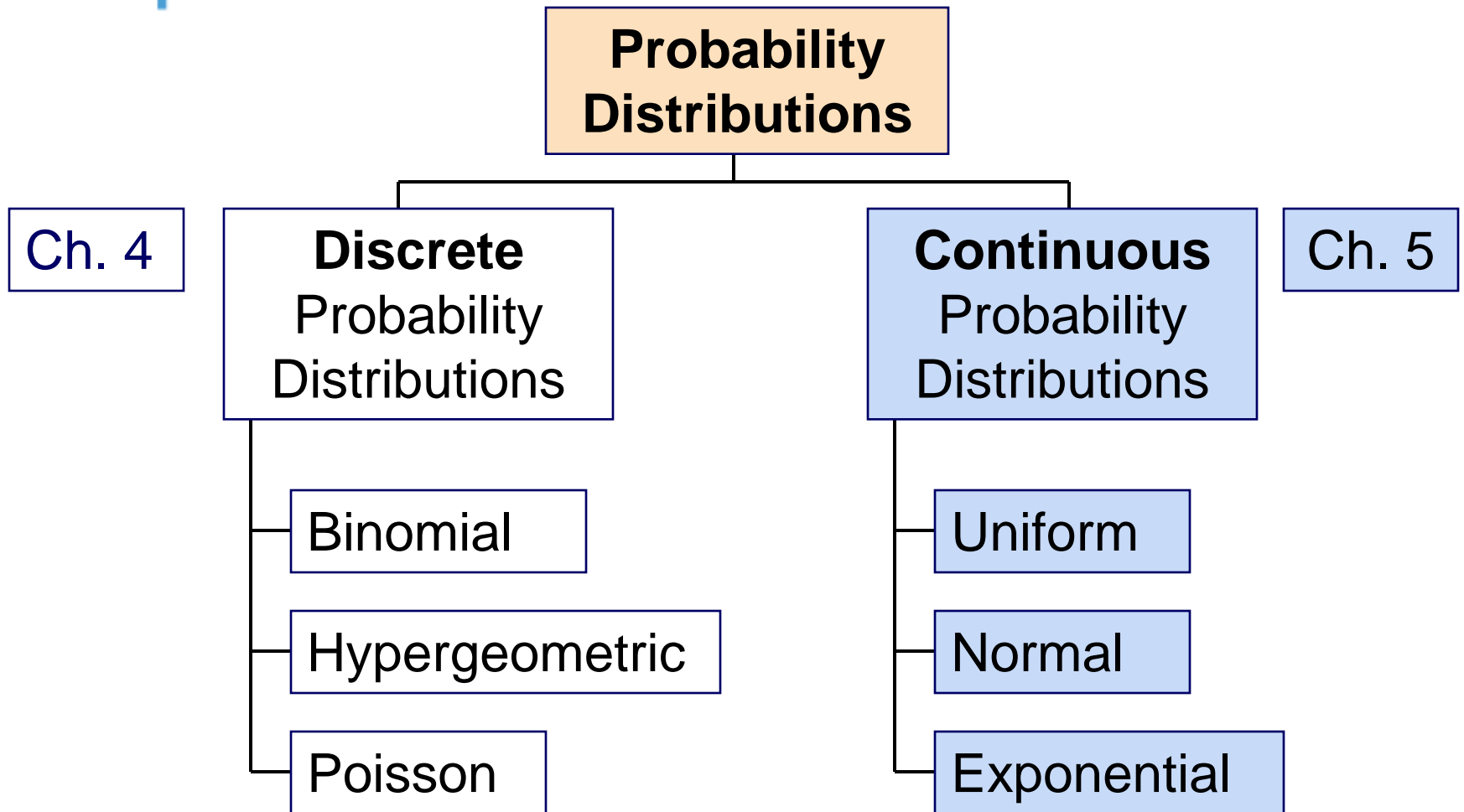
(continued)

After completing this chapter, you should be able to:

- Evaluate the normality assumption
- Use the normal approximation to the binomial distribution
- Recognize when to apply the exponential distribution
- Explain jointly distributed variables and linear combinations of random variables



Probability Distributions



Continuous Probability Distributions

- A **continuous random variable** is a variable that can assume any value in an interval
 - 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.



Cumulative Distribution Function

- The **cumulative distribution function**, $F(x)$, for a continuous random variable X expresses the probability that X does not exceed the value of x

$$F(x) = P(X \leq x)$$

- Let a and b be two possible values of X , with $a < b$. The probability that X lies between a and b is

$$P(a < X < b) = F(b) - F(a)$$



Probability Density Function

The **probability density function**, $f(x)$, of random variable X has the following properties:

1. $f(x) > 0$ for all values of x
2. The area under the probability density function $f(x)$ over all values of the random variable X is equal to 1.0
3. The probability that X lies between two values is the area under the density function graph between the two values



Probability Density Function

(continued)

The **probability density function**, $f(x)$, of random variable X has the following properties:

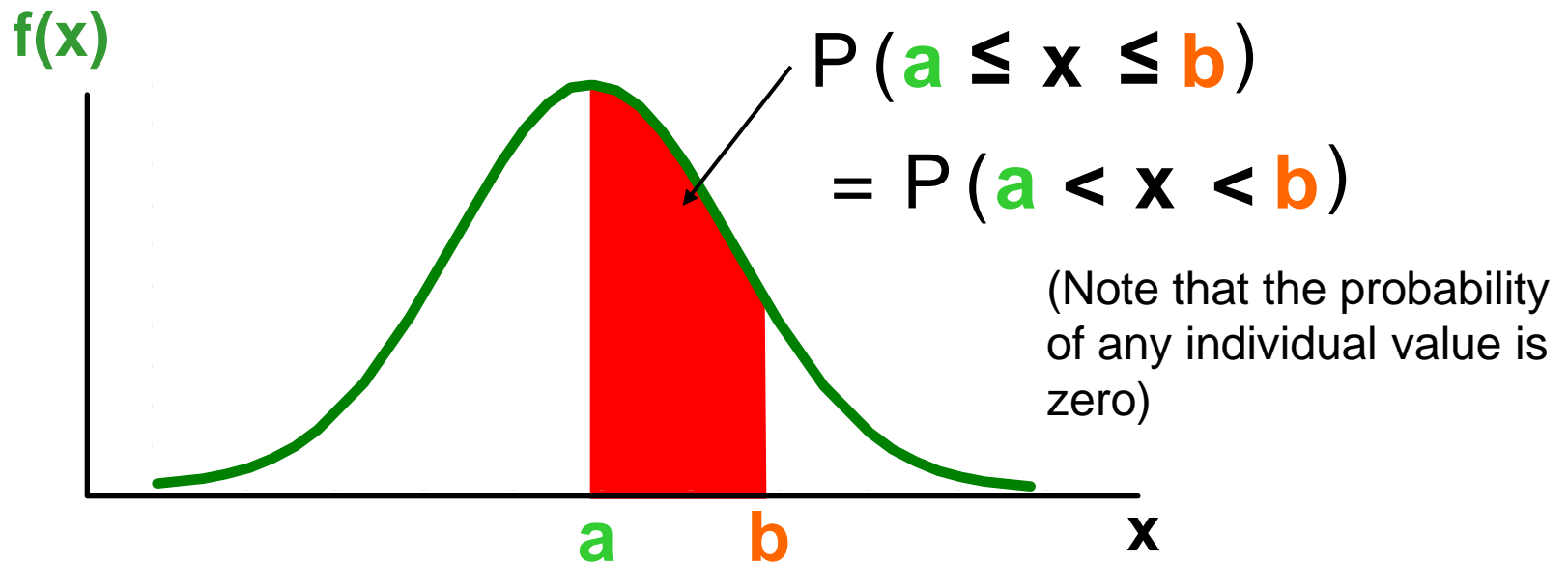
4. The **cumulative density function** $F(x_0)$ is the area under the probability density function $f(x)$ from the minimum x value up to x_0

$$F(x_0) = \int_{x_m}^{x_0} f(x) dx$$

where x_m is the minimum value of the random variable x

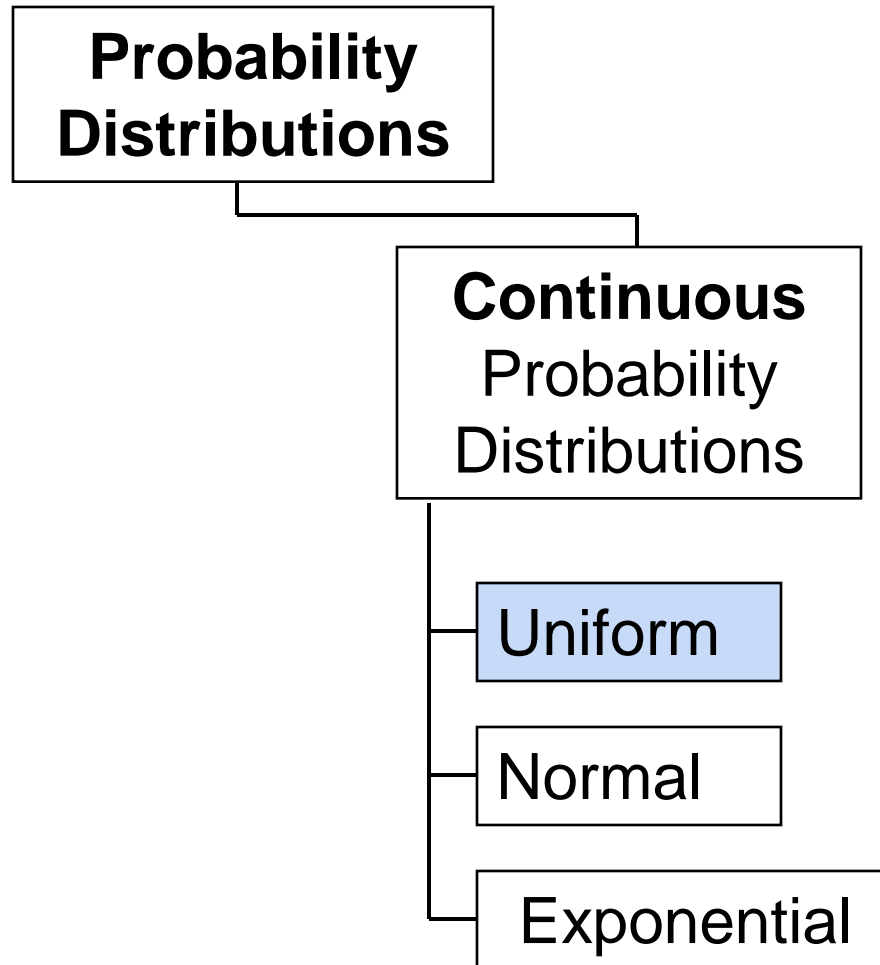
Probability as an Area

Shaded area under the curve is the probability that X is between a and b



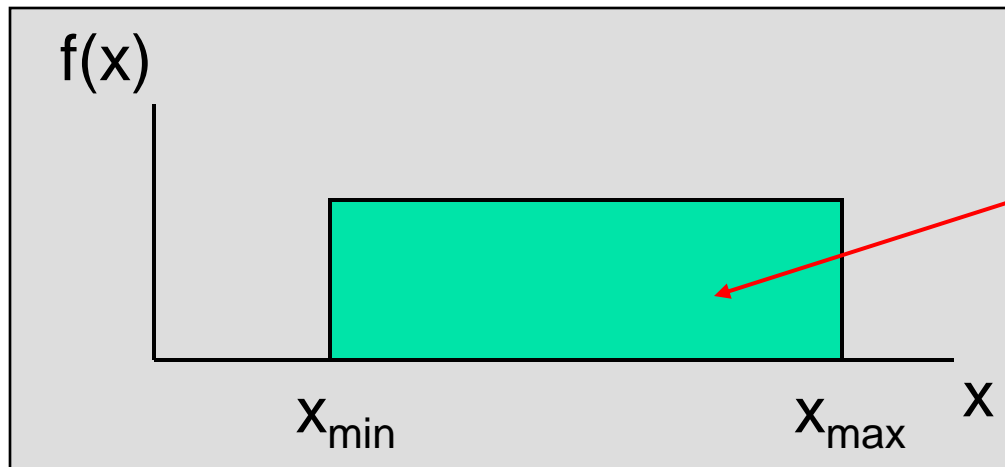


The Uniform Distribution



The Uniform Distribution

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



Total area under the uniform probability density function is 1.0



The Uniform Distribution

(continued)

The Continuous Uniform Distribution:

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

where

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

a = minimum value of x

b = maximum value of x



Properties of the Uniform Distribution

- The **mean** of a uniform distribution is

$$\mu = \frac{a + b}{2}$$

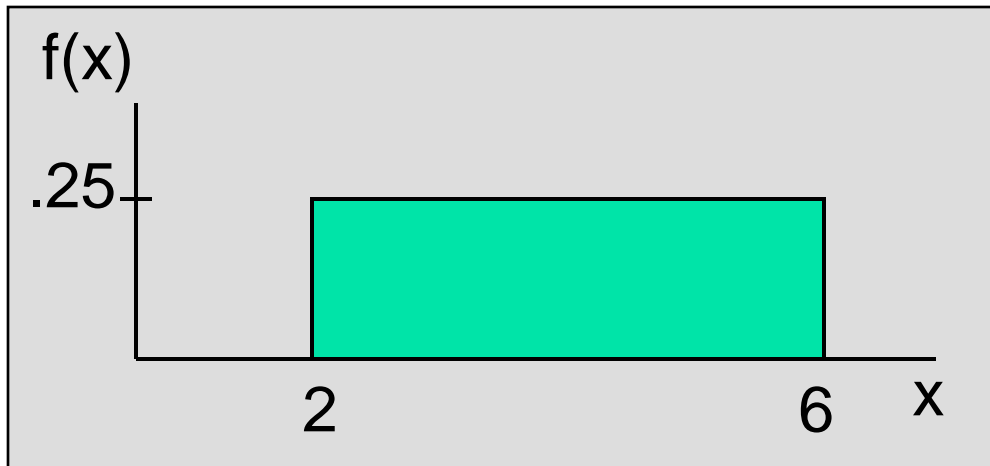
- The **variance** is

$$\sigma^2 = \frac{(b - a)^2}{12}$$

Uniform Distribution Example

Example: Uniform probability distribution over the range $2 \leq x \leq 6$:

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



$$\mu = \frac{a + b}{2} = \frac{2 + 6}{2} = 4$$

$$\sigma^2 = \frac{(b - a)^2}{12} = \frac{(6 - 2)^2}{12} = 1.333$$

Expectations for Continuous Random Variables

- The mean of X , denoted μ_X , is defined as the expected value of X

$$\mu_X = E(X)$$

- The variance of X , denoted σ_X^2 , is defined as the expectation of the squared deviation, $(X - \mu_X)^2$, of a random variable from its mean

$$\sigma_X^2 = E[(X - \mu_X)^2]$$



Linear Functions of Variables

- Let $W = a + bX$, where X has mean μ_X and variance σ_X^2 , and a and b are constants

- Then the mean of W is

$$\mu_W = E(a + bX) = a + b\mu_X$$

- the variance is

$$\sigma_W^2 = \text{Var}(a + bX) = b^2\sigma_X^2$$

- the standard deviation of W is

$$\sigma_W = |b|\sigma_X$$



Linear Functions of Variables

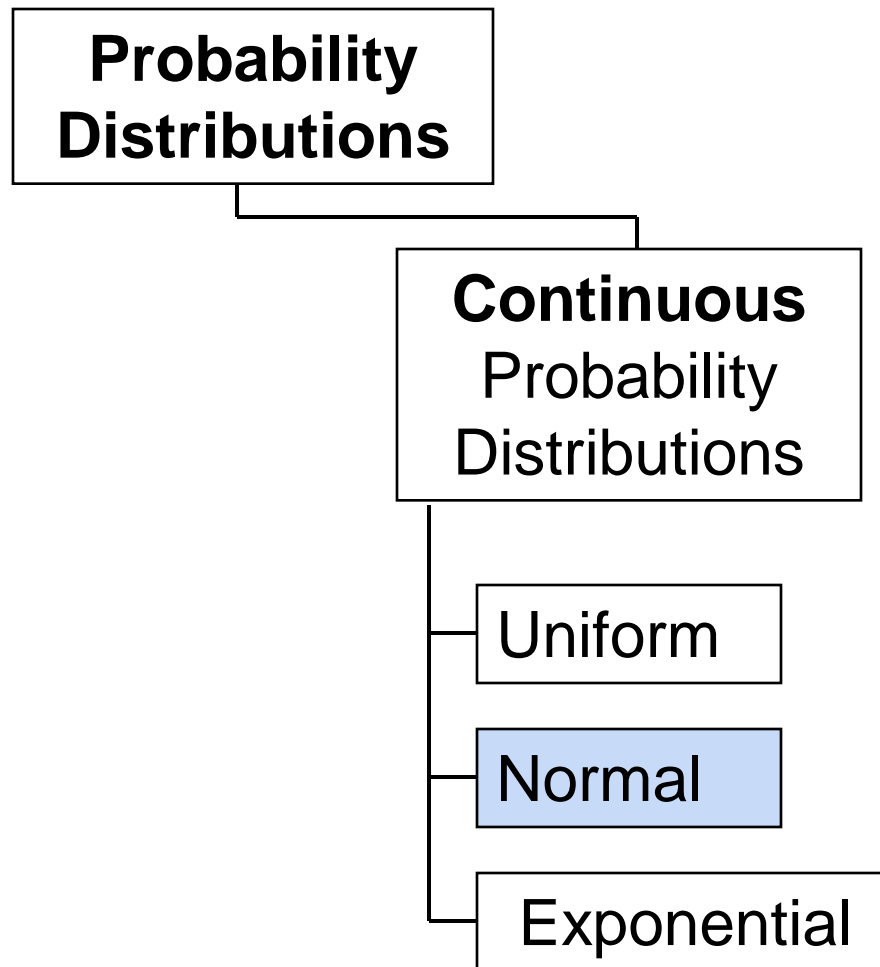
(continued)

- An important special case of the previous results is the standardized random variable

$$Z = \frac{X - \mu_X}{\sigma_X}$$

- which has a mean 0 and variance 1

The Normal Distribution



The Normal Distribution

(continued)

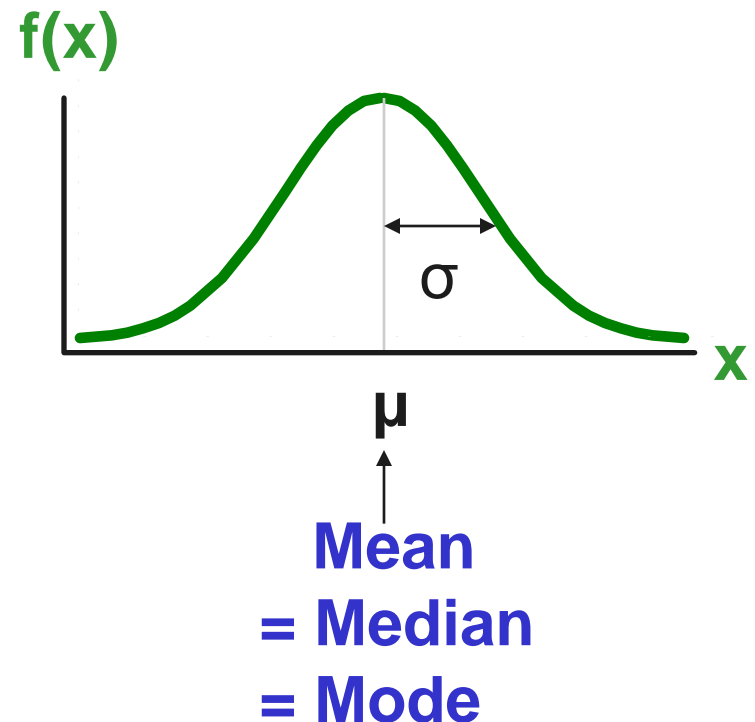
- Bell Shaped
- Symmetrical
- Mean, Median and Mode are Equal

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 $-\infty$



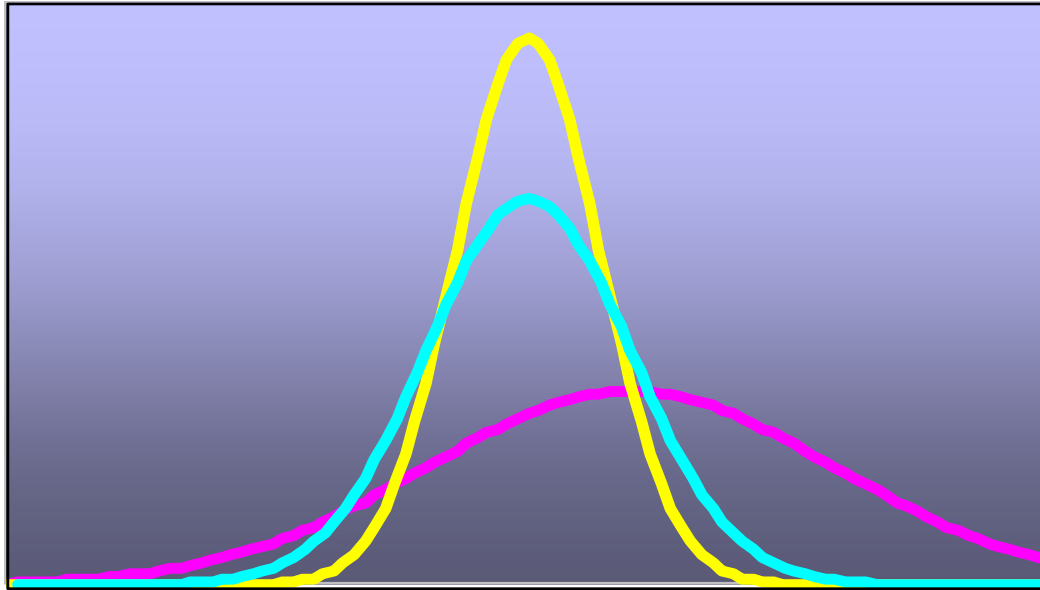


The Normal Distribution

(continued)

- The normal distribution closely approximates the probability distributions of a wide range of random variables
- Distributions of sample means approach a normal distribution given a “large” sample size
- Computations of probabilities are direct and elegant
- The normal probability distribution has led to good business decisions for a number of applications

Many Normal Distributions

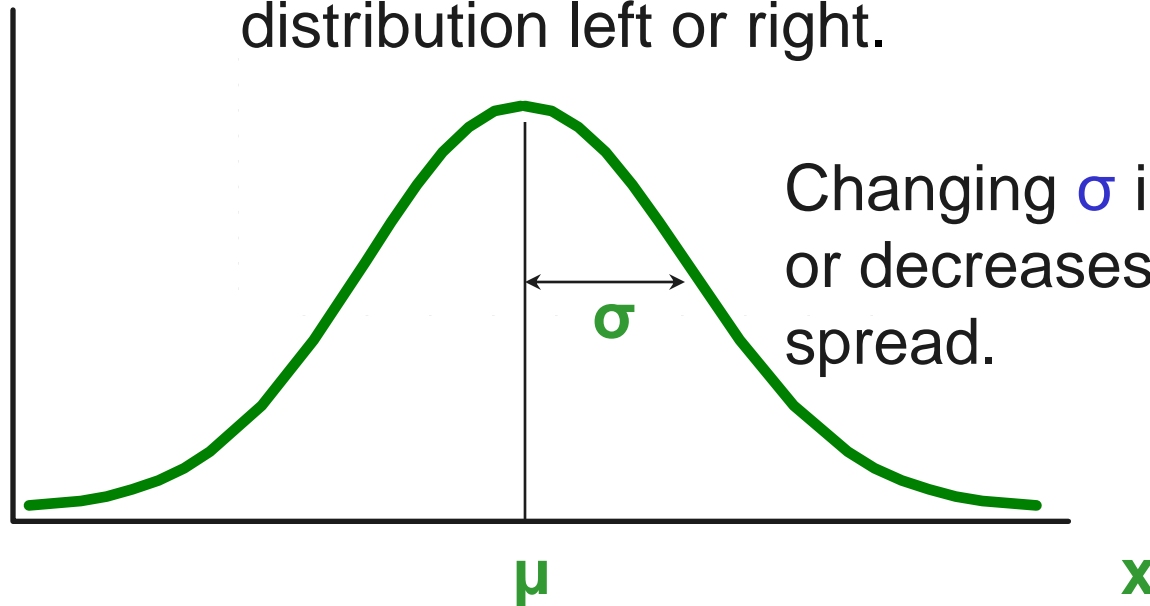


By varying the parameters μ and σ , we obtain different normal distributions

The Normal Distribution Shape

$f(x)$

Changing μ shifts the distribution left or right.



Given the mean μ and variance σ we define the normal distribution using the notation

$$X \sim N(\mu, \sigma^2)$$



The Normal Probability Density Function

- The formula for the normal probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

Where e = the mathematical constant approximated by 2.71828

π = the mathematical constant approximated by 3.14159

μ = the population mean

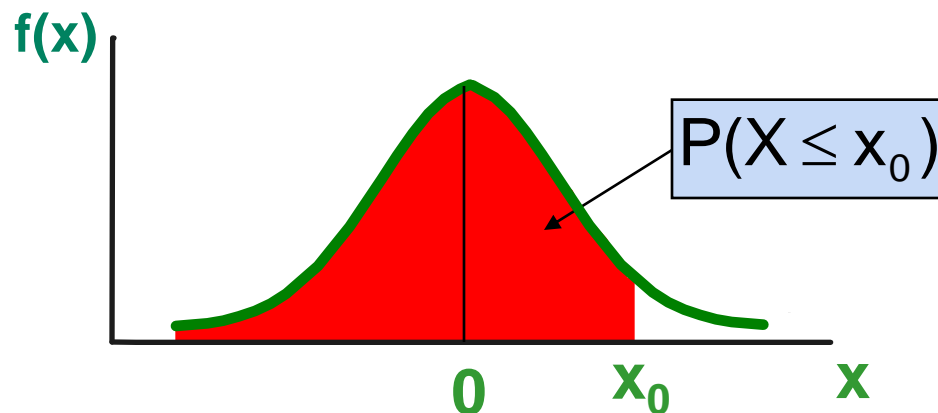
σ = the population standard deviation

x = any value of the continuous variable, $-\infty < x < \infty$

Cumulative Normal Distribution

- For a normal random variable X with mean μ and variance σ^2 , i.e., $X \sim N(\mu, \sigma^2)$, the **cumulative distribution function** is

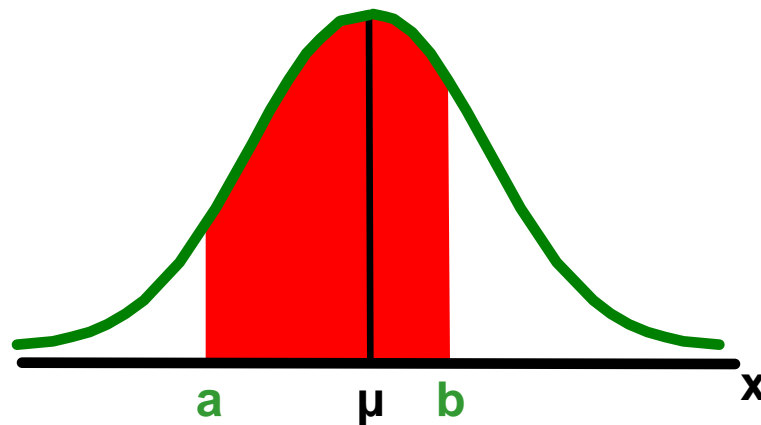
$$F(x_0) = P(X \leq x_0)$$



Finding Normal Probabilities

The probability for a range of values is measured by the area under the curve

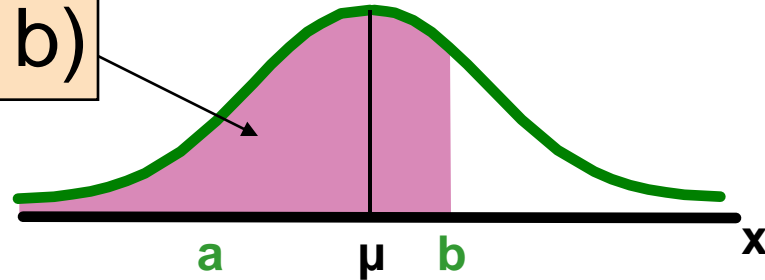
$$P(a < X < b) = F(b) - F(a)$$



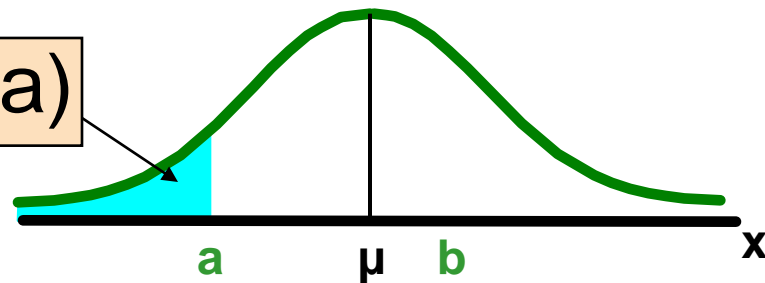
Finding Normal Probabilities

(continued)

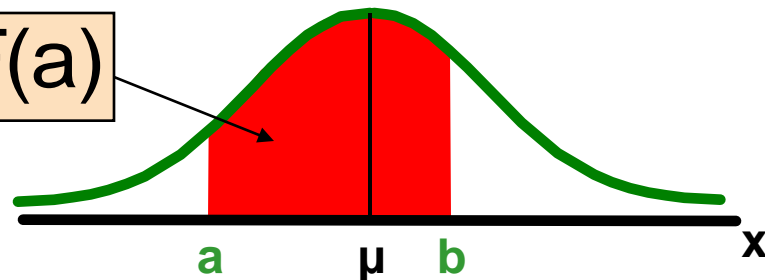
$$F(b) = P(X < b)$$



$$F(a) = P(X < a)$$



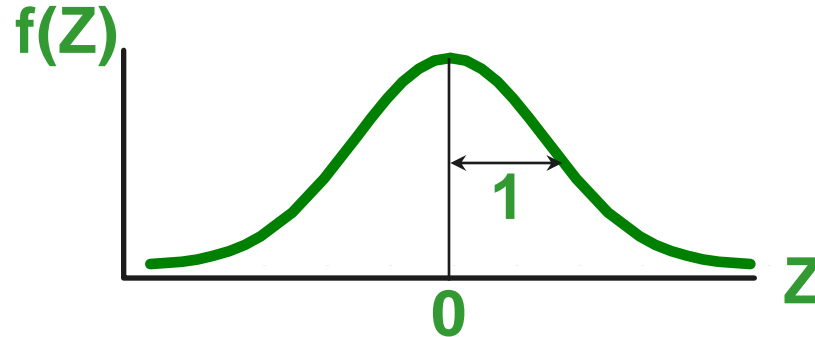
$$P(a < X < b) = F(b) - F(a)$$



The Standardized Normal

- Any normal distribution (with any mean and variance combination) can be transformed into the standardized normal distribution (Z), with mean 0 and variance 1

$$Z \sim N(0,1)$$



- Need to transform X units into Z units by subtracting the mean of X and dividing by its standard deviation

$$Z = \frac{X - \mu}{\sigma}$$



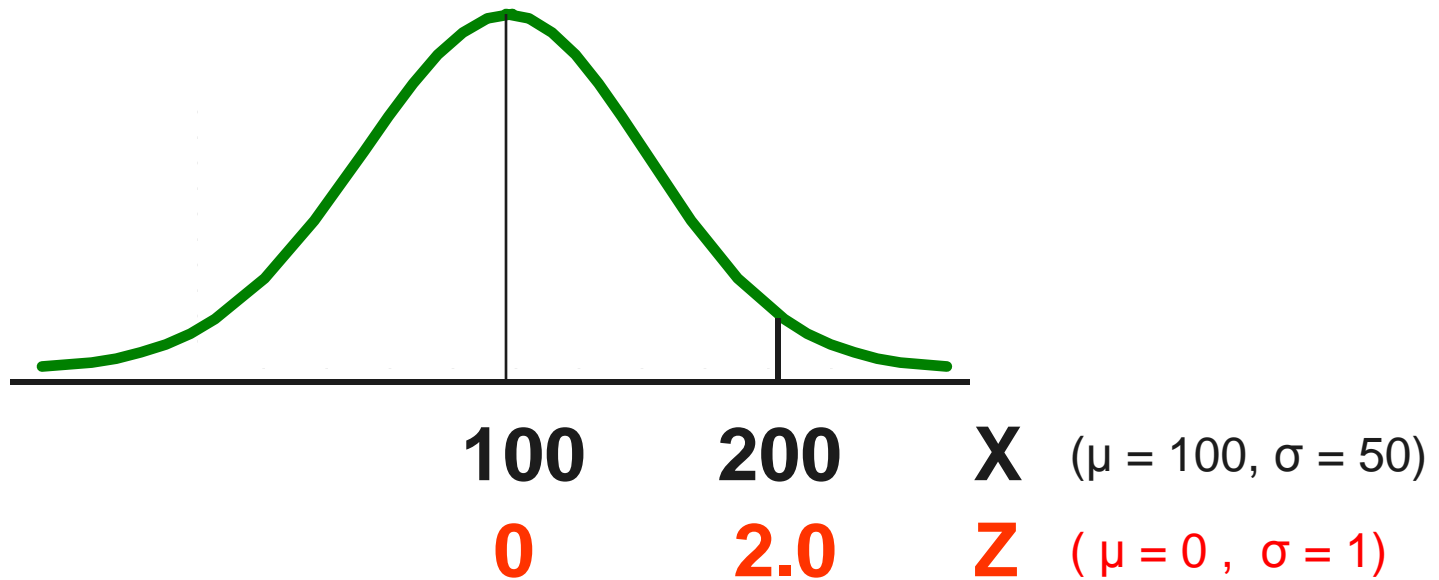
Example

- If X is distributed normally with mean of 100 and standard deviation of 50, the Z value for $X = 200$ is

$$Z = \frac{X - \mu}{\sigma} = \frac{200 - 100}{50} = 2.0$$

- This says that $X = 200$ is two standard deviations (2 increments of 50 units) above the mean of 100.

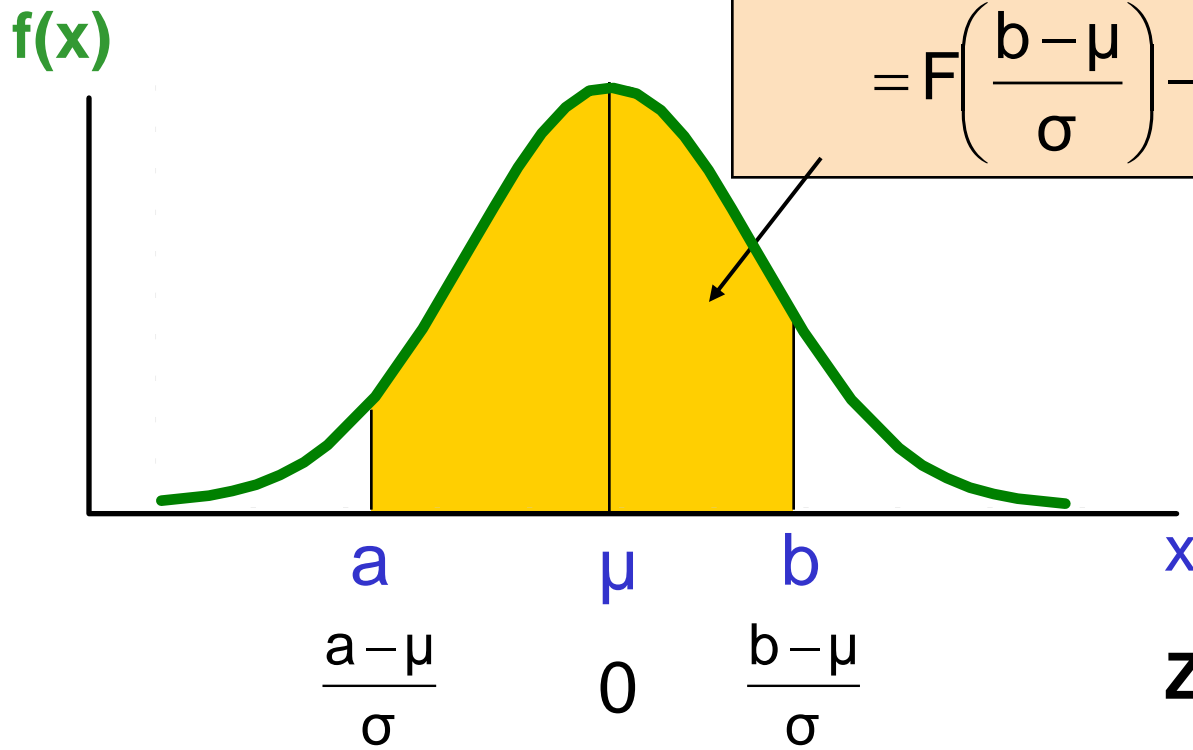
Comparing X and Z units



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)

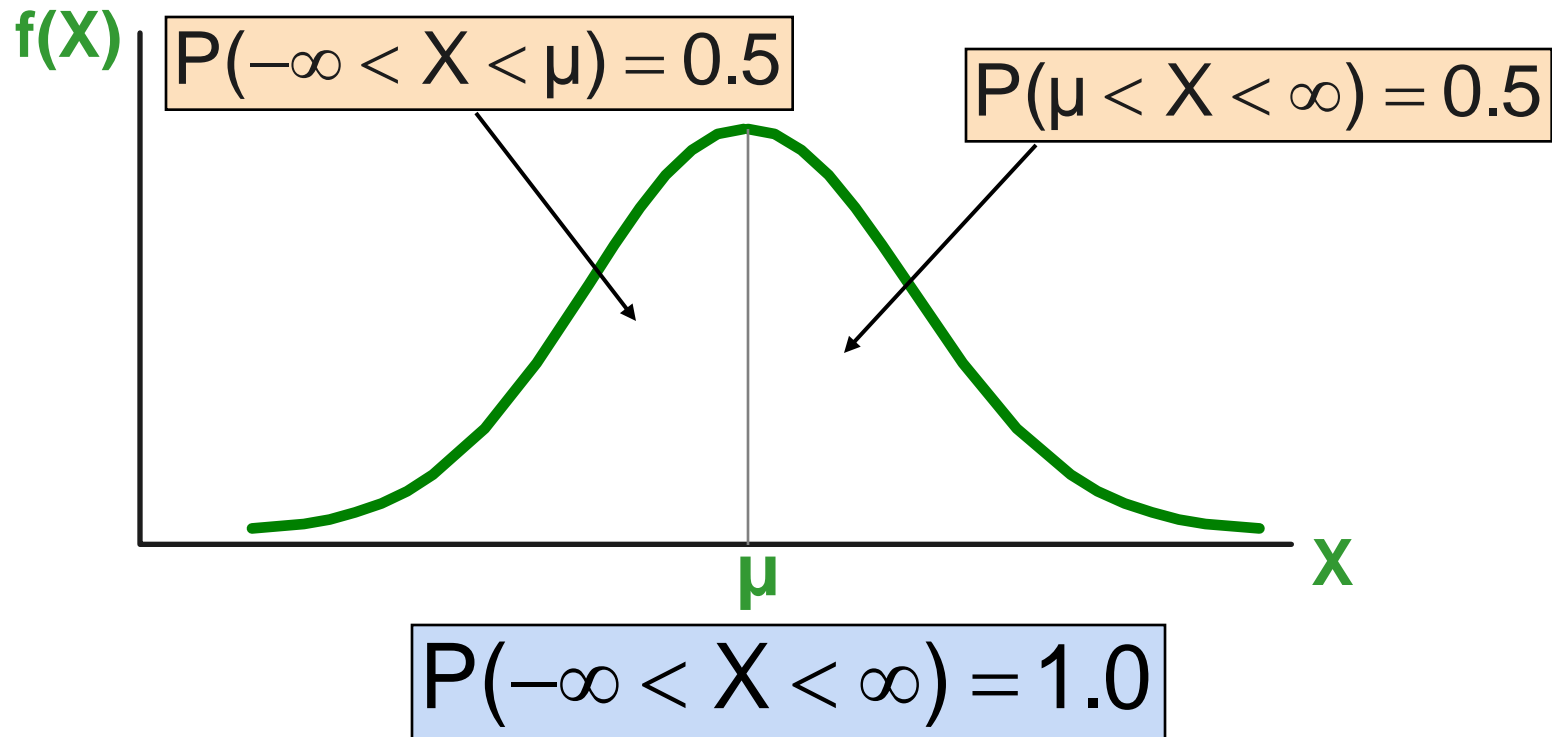
Finding Normal Probabilities

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right) \\ &= F\left(\frac{b - \mu}{\sigma}\right) - F\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$



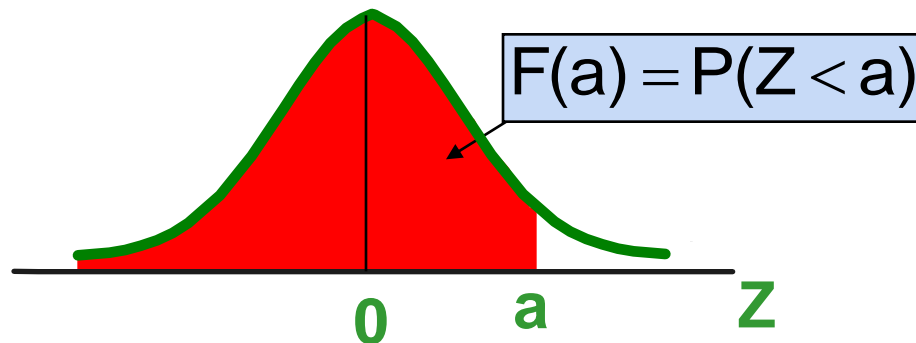
Probability as Area Under the Curve

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



Appendix Table 1

- The Standardized Normal table in the textbook (Appendix Table 1) shows values of the cumulative normal distribution function
- For a given Z-value a , the table shows $F(a)$ (the area under the curve from negative infinity to a)

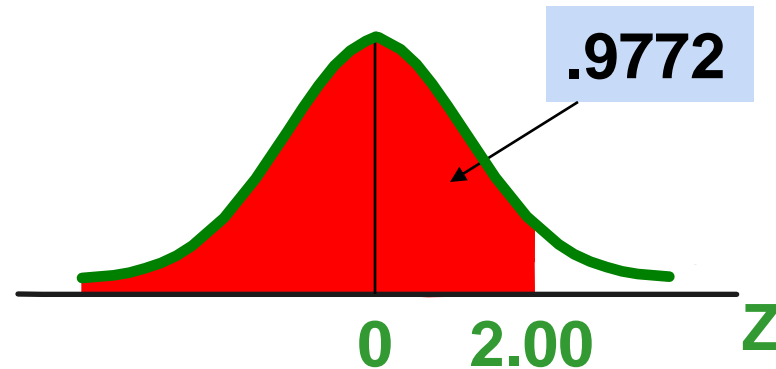


The Standardized Normal Table

- Appendix Table 1 gives the probability $F(a)$ for any value a

Example:

$$P(Z < 2.00) = .9772$$



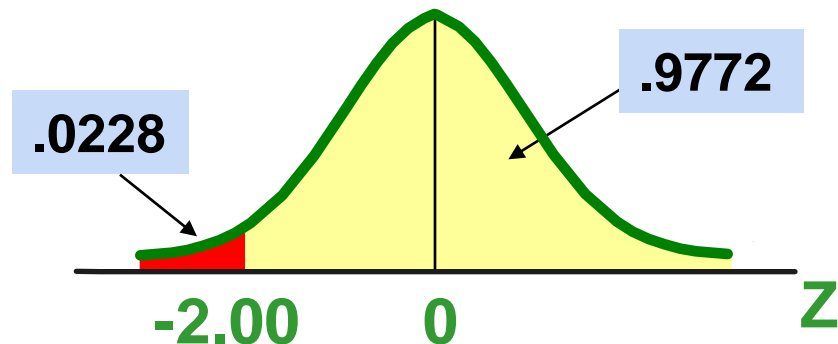
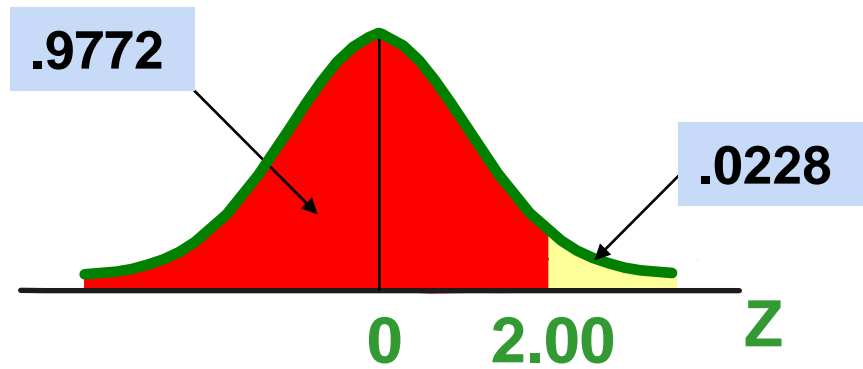
The Standardized Normal Table

(continued)

- For **negative Z-values**, use the fact that the distribution is symmetric to find the needed probability:

Example:

$$\begin{aligned} P(Z < -2.00) &= 1 - 0.9772 \\ &= 0.0228 \end{aligned}$$





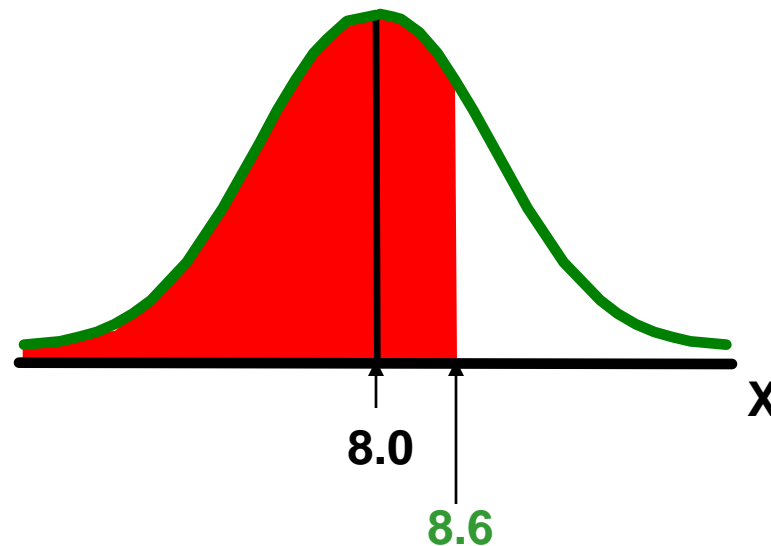
General Procedure for Finding Probabilities

To find $P(a < X < b)$ when X is distributed normally:

- Draw the normal curve for the problem in terms of X
- Translate X -values to Z -values
- Use the Cumulative Normal Table

Finding Normal Probabilities

- Suppose X is normal with mean 8.0 and standard deviation 5.0
- Find $P(X < 8.6)$

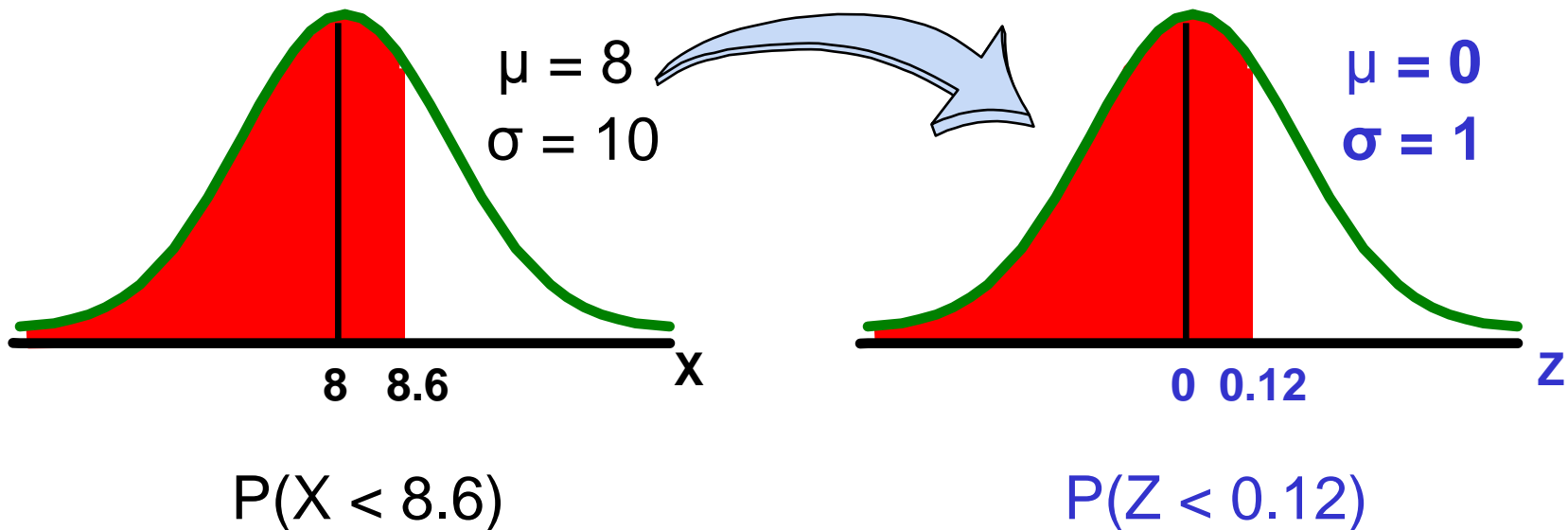


Finding Normal Probabilities

(continued)

- Suppose X is normal with mean 8.0 and standard deviation 5.0. Find $P(X < 8.6)$

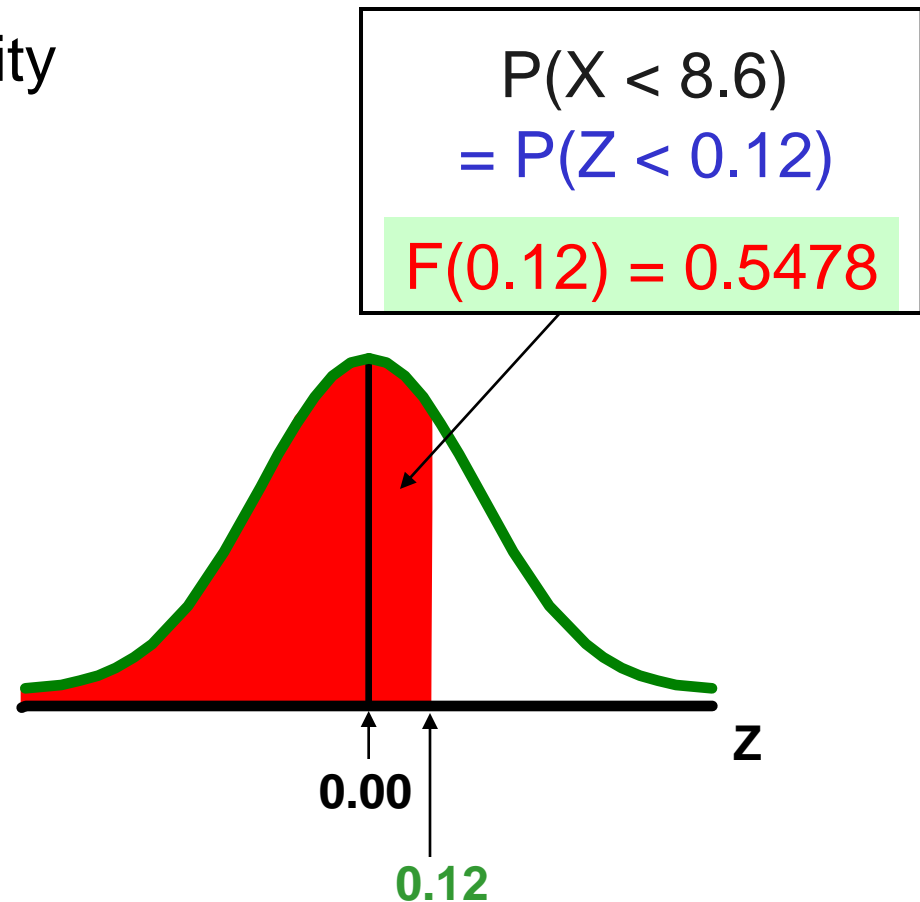
$$Z = \frac{X - \mu}{\sigma} = \frac{8.6 - 8.0}{5.0} = 0.12$$



Solution: Finding $P(Z < 0.12)$

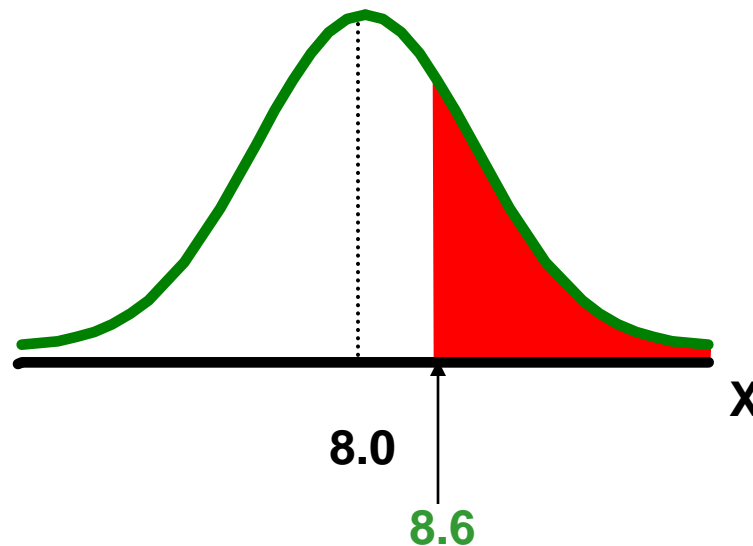
Standardized Normal Probability Table (Portion)

z	F(z)
.10	.5398
.11	.5438
.12	.5478
.13	.5517



Upper Tail Probabilities

- Suppose X is normal with mean 8.0 and standard deviation 5.0.
- Now Find $P(X > 8.6)$



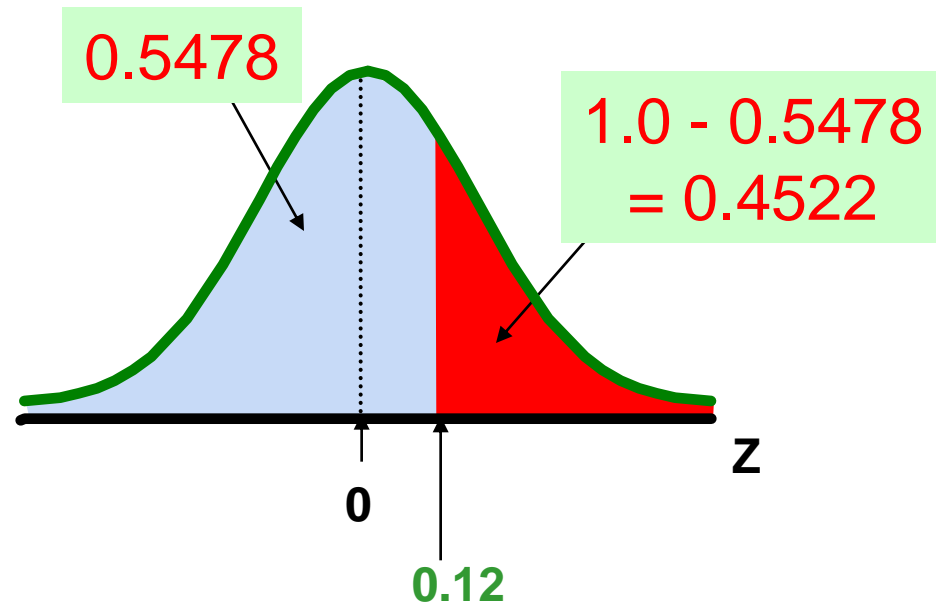
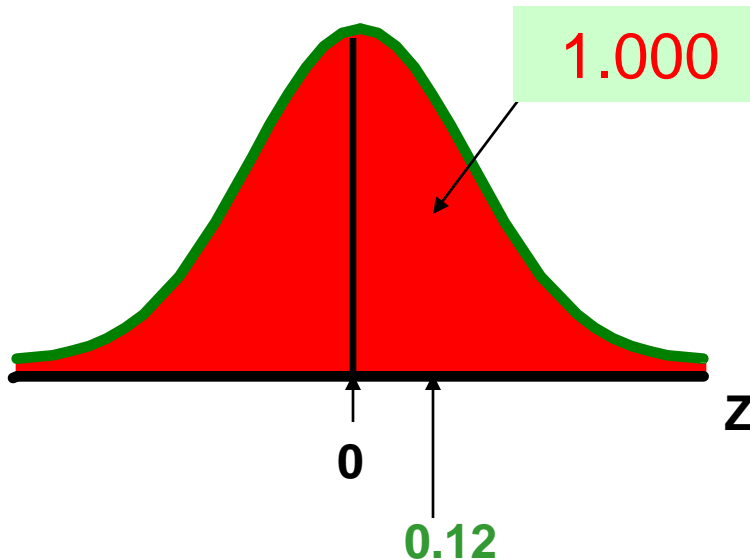
Upper Tail Probabilities

(continued)

- Now Find $P(X > 8.6)$...

$$P(X > 8.6) = P(Z > 0.12) = 1.0 - P(Z \leq 0.12)$$

$$= 1.0 - 0.5478 = 0.4522$$





Finding the X value for a Known Probability

- Steps to find the X value for a known probability:
 1. Find the Z value for the known probability
 2. Convert to X units using the formula:

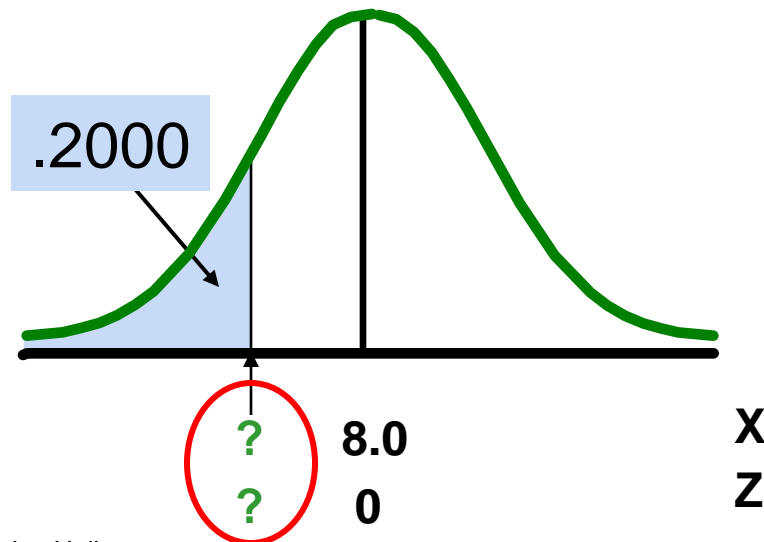
$$X = \mu + Z\sigma$$

Finding the X value for a Known Probability

(continued)

Example:

- Suppose X is normal with mean 8.0 and standard deviation 5.0.
- Now find the X value so that only 20% of all values are below this X



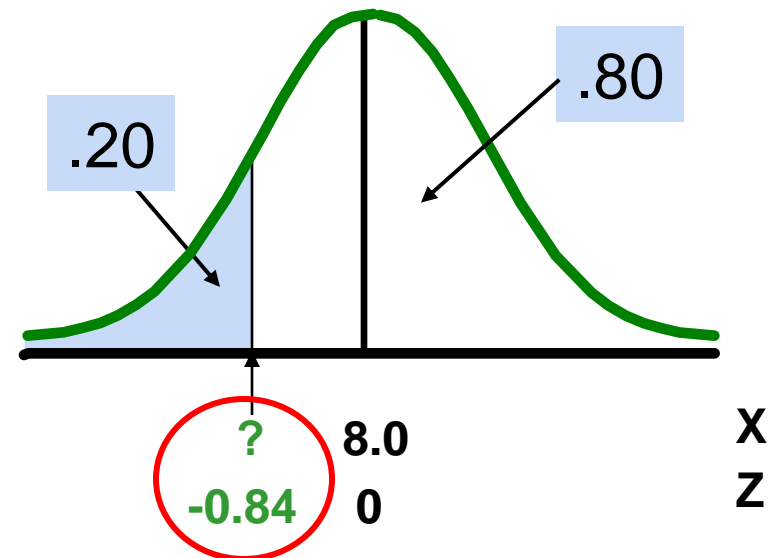
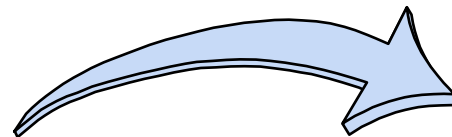
Find the Z value for 20% in the Lower Tail

1. Find the Z value for the known probability

Standardized Normal Probability Table (Portion)

z	F(z)
.82	.7939
.83	.7967
.84	.7995
.85	.8023

- 20% area in the lower tail is consistent with a Z value of **-0.84**





Finding the X value

2. Convert to X units using the formula:

$$\begin{aligned} X &= \mu + Z\sigma \\ &= 8.0 + (-0.84)5.0 \\ &= 3.80 \end{aligned}$$

So 20% of the values from a distribution with mean 8.0 and standard deviation 5.0 are less than 3.80



Assessing Normality

- Not all continuous random variables are normally distributed
- It is important to evaluate how well the data is approximated by a normal distribution



The Normal Probability Plot

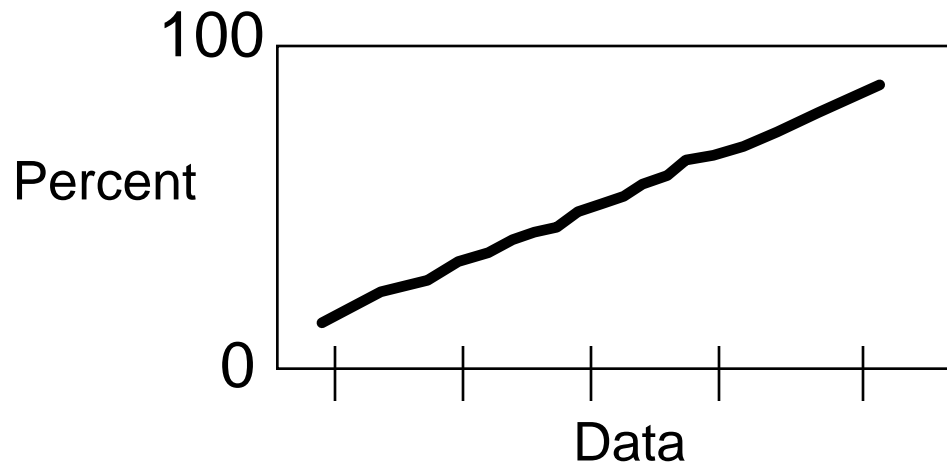
- Normal probability plot
 - Arrange data from low to high values
 - Find cumulative normal probabilities for all values
 - Examine a plot of the observed values vs. cumulative probabilities (with the cumulative normal probability on the vertical axis and the observed data values on the horizontal axis)
 - Evaluate the plot for evidence of linearity



The Normal Probability Plot

(continued)

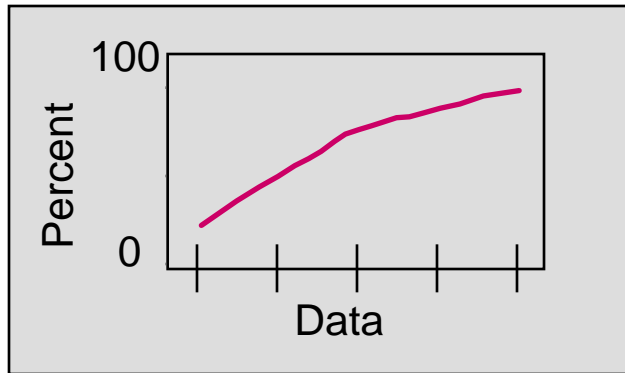
A normal probability plot for data from a normal distribution will be **approximately linear**:



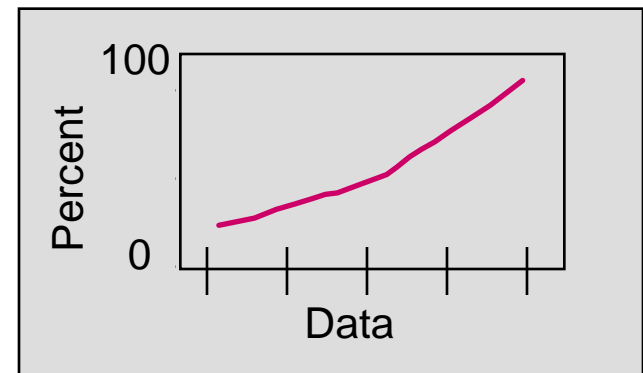
The Normal Probability Plot

(continued)

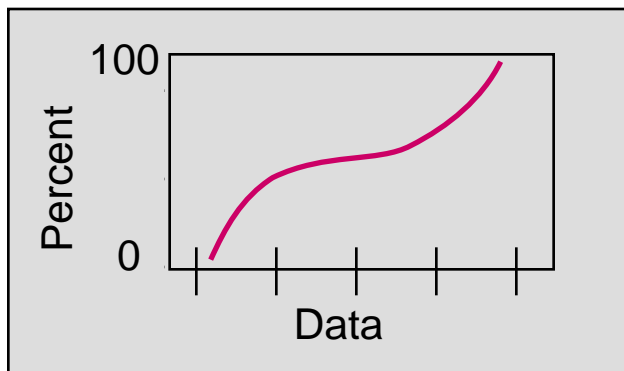
Left-Skewed



Right-Skewed



Uniform



Nonlinear plots indicate a deviation from normality

Normal Distribution Approximation for Binomial Distribution

- Recall the binomial distribution:
 - n independent trials
 - probability of success on any given trial = P
- Random variable X:
 - $X_i = 1$ if the i^{th} trial is “success”
 - $X_i = 0$ if the i^{th} trial is “failure”

$$E(X) = \mu = nP$$

$$\text{Var}(X) = \sigma^2 = nP(1-P)$$



Normal Distribution Approximation for Binomial Distribution

(continued)

- The shape of the binomial distribution is **approximately normal** if n is large
- The normal is a good approximation to the binomial when $nP(1 - P) > 5$
- Standardize to Z from a binomial distribution:

$$Z = \frac{X - E(X)}{\sqrt{\text{Var}(X)}} = \frac{X - np}{\sqrt{nP(1-P)}}$$



Normal Distribution Approximation for Binomial Distribution

(continued)

- Let X be the number of successes from n independent trials, each with probability of success P .
- If $nP(1 - P) > 5$,

$$P(a < X < b) = P\left(\frac{a - nP}{\sqrt{nP(1-P)}} \leq Z \leq \frac{b - nP}{\sqrt{nP(1-P)}}\right)$$

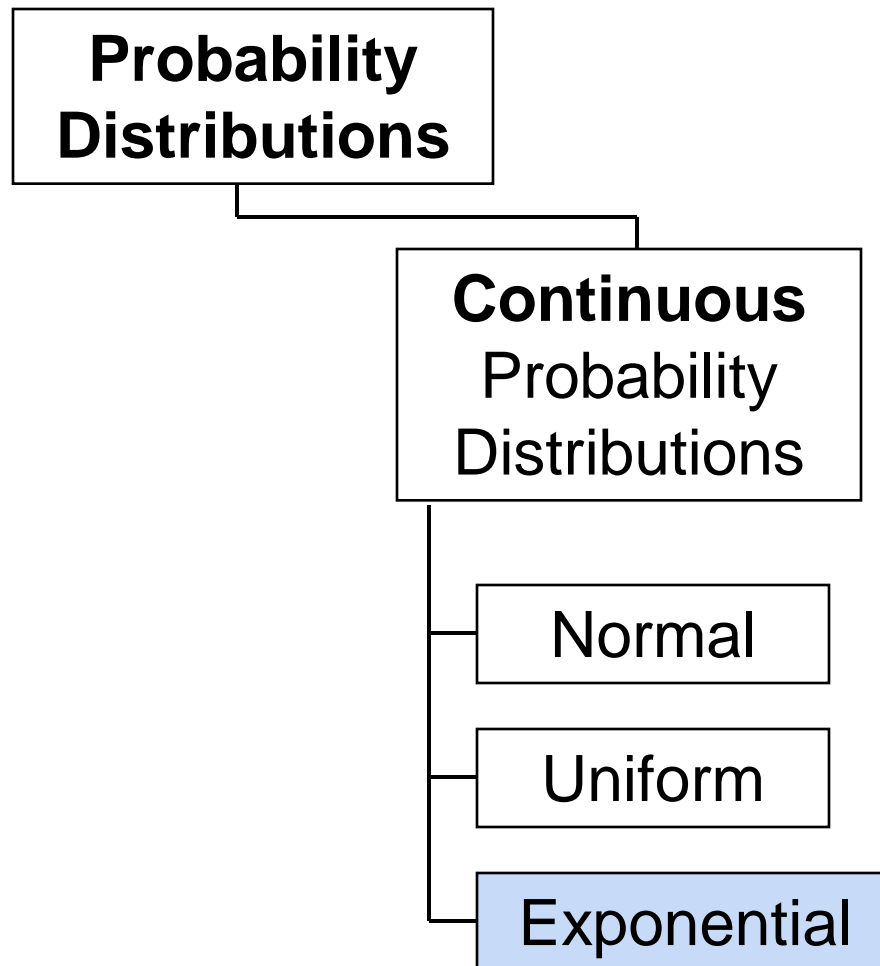


Binomial Approximation Example

- 40% of all voters support ballot proposition A. What is the probability that between 76 and 80 voters indicate support in a sample of $n = 200$?
 - $E(X) = \mu = nP = 200(0.40) = 80$
 - $\text{Var}(X) = \sigma^2 = nP(1 - P) = 200(0.40)(1 - 0.40) = 48$
(note: $nP(1 - P) = 48 > 5$)

$$\begin{aligned} P(76 < X < 80) &= P\left(\frac{76 - 80}{\sqrt{200(0.4)(1-0.4)}} \leq Z \leq \frac{80 - 80}{\sqrt{200(0.4)(1-0.4)}}\right) \\ &= P(-0.58 < Z < 0) \\ &= F(0) - F(-0.58) \\ &= 0.5000 - 0.2810 = 0.2190 \end{aligned}$$

The Exponential Distribution





The Exponential Distribution

- Used to model the **length of time 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



The Exponential Distribution

(continued)

- The exponential random variable T ($t > 0$) has a probability density function

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

- Where
 - λ is the mean number of occurrences per unit time
 - t is the number of time units until the next occurrence
 - $e = 2.71828$
- T is said to follow an exponential probability distribution



The Exponential Distribution

- Defined by a single parameter, its mean λ (lambda)
- The **cumulative distribution function** (the probability that an arrival time is less than some specified time t) is

$$F(t) = 1 - e^{-\lambda t}$$

where e = mathematical constant approximated by 2.71828
 λ = the population mean number of arrivals per unit
 t = any value of the continuous variable where $t > 0$



Exponential Distribution Example

Example: Customers arrive at the service counter at the rate of 15 per hour. What is the probability that the arrival time between consecutive customers is less than three minutes?

- The mean number of arrivals per hour is 15, so $\lambda = 15$
- Three minutes is .05 hours
- $P(\text{arrival time} < .05) = 1 - e^{-\lambda X} = 1 - e^{-(15)(.05)} = 0.5276$
- So there is a 52.76% probability that the arrival time between successive customers is less than three minutes

Joint Cumulative Distribution Functions

- Let X_1, X_2, \dots, X_k be continuous random variables
- Their **joint cumulative distribution function**,

$$F(x_1, x_2, \dots, x_k)$$

defines the probability that simultaneously X_1 is less than x_1 , X_2 is less than x_2 , and so on; that is

$$F(x_1, x_2, \dots, x_k) = P(X_1 < x_1 \cap X_2 < x_2 \cap \dots \cap X_k < x_k)$$



Joint Cumulative Distribution Functions

(continued)

- The cumulative distribution functions

$$F(x_1), F(x_2), \dots, F(x_k)$$

of the individual random variables are called their **marginal distribution functions**

- The random variables are **independent** if and only if

$$F(x_1, x_2, \dots, x_k) = F(x_1)F(x_2) \cdots F(x_k)$$



Covariance

- Let X and Y be continuous random variables, with means μ_x and μ_y
- The expected value of $(X - \mu_x)(Y - \mu_y)$ is called the **covariance** between X and Y

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

- An alternative but equivalent expression is

$$\text{Cov}(X, Y) = E(XY) - \mu_x\mu_y$$

- If the random variables X and Y are independent, then the covariance between them is 0. However, the converse is not true.



Correlation

- Let X and Y be jointly distributed random variables.
- The **correlation** between X and Y is

$$\rho = \text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$



Sums of Random Variables

Let X_1, X_2, \dots, X_k be k random variables with means $\mu_1, \mu_2, \dots, \mu_k$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$. Then:

- The mean of their sum is the sum of their means

$$E(X_1 + X_2 + \dots + X_k) = \mu_1 + \mu_2 + \dots + \mu_k$$



Sums of Random Variables

(continued)

Let X_1, X_2, \dots, X_k be k random variables with means $\mu_1, \mu_2, \dots, \mu_k$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$. Then:

- If the covariance between every pair of these random variables is 0, then the variance of their sum is the sum of their variances

$$\text{Var}(X_1 + X_2 + \dots + X_k) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2$$

- However, if the covariances between pairs of random variables are not 0, the variance of their sum is

$$\text{Var}(X_1 + X_2 + \dots + X_k) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2 + 2 \sum_{i=1}^{K-1} \sum_{j=i+1}^K \text{Cov}(X_i, X_j)$$



Differences Between Two Random Variables

For two random variables, X and Y

- The mean of their difference is the difference of their means; that is

$$E(X - Y) = \mu_X - \mu_Y$$

- If the covariance between X and Y is 0, then the variance of their difference is

$$\text{Var}(X - Y) = \sigma_X^2 + \sigma_Y^2$$

- If the covariance between X and Y is not 0, then the variance of their difference is

$$\text{Var}(X - Y) = \sigma_X^2 + \sigma_Y^2 - 2\text{Cov}(X, Y)$$



Linear Combinations of Random Variables

- A linear combination of two random variables, X and Y , (where a and b are constants) is

$$W = aX + bY$$

- The mean of W is

$$\mu_W = E[W] = E[aX + bY] = a\mu_X + b\mu_Y$$



Linear Combinations of Random Variables

(continued)

- The variance of W is

$$\sigma_W^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\text{Cov}(X, Y)$$

- Or using the correlation,

$$\sigma_W^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\text{Corr}(X, Y)\sigma_X\sigma_Y$$

- If both X and Y are joint normally distributed random variables then the linear combination, W , is also normally distributed



Example

- Two tasks must be performed by the same worker.
 - X = minutes to complete task 1; $\mu_x = 20$, $\sigma_x = 5$
 - Y = minutes to complete task 2; $\mu_y = 20$, $\sigma_y = 5$
 - X and Y are normally distributed and independent
- What is the mean and standard deviation of the time to complete both tasks?



Example

(continued)

- X = minutes to complete task 1; $\mu_x = 20$, $\sigma_x = 5$
- Y = minutes to complete task 2; $\mu_y = 30$, $\sigma_y = 8$
- What are the mean and standard deviation for the time to complete both tasks?

$$W = X + Y$$

$$\mu_W = \mu_X + \mu_Y = 20 + 30 = 50$$

- Since X and Y are independent, $\text{Cov}(X, Y) = 0$, so

$$\sigma_W^2 = \sigma_X^2 + \sigma_Y^2 + 2\text{Cov}(X, Y) = (5)^2 + (8)^2 = 89$$

- The standard deviation is

$$\sigma_W = \sqrt{89} = 9.434$$



Portfolio Analysis

- A financial **portfolio** can be viewed as a linear combination of separate financial instruments

$$\begin{aligned} \left(\begin{array}{c} \text{Return on} \\ \text{portfolio} \end{array} \right) &= \left(\begin{array}{c} \text{Proportion of} \\ \text{portfolio value} \\ \text{in stock 1} \end{array} \right) \times \left(\begin{array}{c} \text{Stock 1} \\ \text{return} \end{array} \right) + \left(\begin{array}{c} \text{Proportion of} \\ \text{portfolio value} \\ \text{in stock 2} \end{array} \right) \times \left(\begin{array}{c} \text{Stock 2} \\ \text{return} \end{array} \right) \\ &\dots + \left(\begin{array}{c} \text{Proportion of} \\ \text{portfolio value} \\ \text{in stock N} \end{array} \right) \times \left(\begin{array}{c} \text{Stock N} \\ \text{return} \end{array} \right) \end{aligned}$$



Portfolio Analysis Example

- Consider two stocks, A and B
 - The price of Stock A is normally distributed with mean 12 and variance 4
 - The price of Stock B is normally distributed with mean 20 and variance 16
 - The stock prices have a positive correlation, $\rho_{AB} = .50$
- Suppose you own
 - 10 shares of Stock A
 - 30 shares of Stock B



Portfolio Analysis Example

(continued)

- The mean and variance of this stock portfolio are: (Let W denote the distribution of portfolio value)

$$\mu_W = 10\mu_A + 20\mu_B = (10)(12) + (30)(20) = 720$$

$$\begin{aligned}\sigma_W^2 &= 10^2 \sigma_A^2 + 30^2 \sigma_B^2 + (2)(10)(30)\text{Corr}(A,B)\sigma_A \sigma_B \\ &= 10^2 (4)^2 + 30^2 (16)^2 + (2)(10)(30)(.50)(4)(16) \\ &= 251,200\end{aligned}$$



Portfolio Analysis Example

(continued)

- What is the probability that your portfolio value is less than \$500?

$$\mu_W = 720$$

$$\sigma_W = \sqrt{251,200} = 501.20$$

- The Z value for 500 is

$$Z = \frac{500 - 720}{501.20} = -0.44$$

- $P(Z < -0.44) = 0.3300$

- So the probability is 0.33 that your portfolio value is less than \$500.



Chapter Summary

- Defined continuous random variables
- Presented key continuous probability distributions and their properties
 - uniform, normal, exponential
- Found probabilities using formulas and tables
- Interpreted normal probability plots
- Examined when to apply different distributions
- Applied the normal approximation to the binomial distribution
- Reviewed properties of jointly distributed continuous random variables