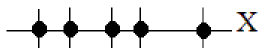
CHAPTERS 5 & 6: CONTINUOUS RANDOM VARIABLES

DISCRETE RANDOM VARIABLE:	CONTINUOUS RANDOM VARIABLE:
Variable can take on only certain specified values. There are gaps between possible data values. Values may be "counting numbers" or may be a collection of numbers from the context of the situation. 	Variable can take on all numbers in a specific interval of values. There are no gaps or breaks between possible data values. 

The information below on this page is adapted from Introductory Statistics from OpenStax available for download for free http://cnx.org/contents/30189442-6998-4686-ac05-ed152b91b9de@17.43:34/Introductory_Statistics

Continuous probability distribution: Graph is a continuous curve.

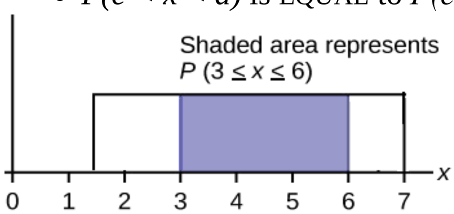
- **Curve** is called the **probability density function** (abbreviated **pdf**).
- Symbol $f(x)$ represents the curve; $y = f(x)$ is the function that draws the graph

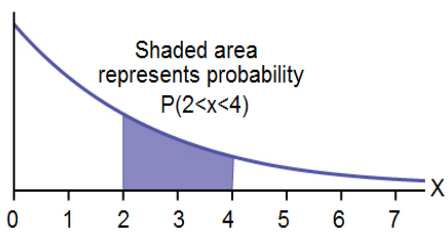
Probability is measured for intervals of x values.

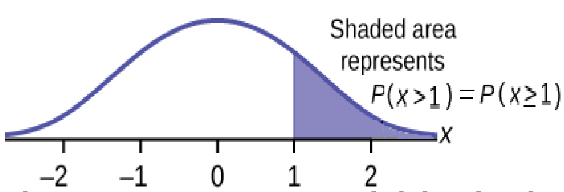
- ***P(c < x < d)* and *P(c ≤ x ≤ d)* represent: probability that random variable *X* is in the interval between values *x* = *c* and *x* = *d*.**

Probability is represented by area under the pdf curve.

- **Probability (area under the curve) is calculated by** a different function called the **cumulative distribution function** (abbreviated **cdf**).
- The entire area under the curve and above the x-axis is equal to one.
- ***P(c < x < d)* and *P(c ≤ x ≤ d)* is equal to the area that is:**
  - **under the pdf curve f(x) and above the *x*-axis**
  - **between *x* = *c* and *x* = *d***
- ***P(x = c)* = 0** : probability that *x* takes on any single individual value is zero:
- *P(c < x < d)* is EQUAL to *P(c ≤ x ≤ d)* because both have the same area for their probabilities



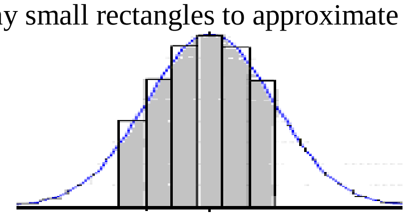




We find areas for probability by using:

- geometry
- formulas
- technology
- probability tables.

The formulas or technology we will use results that come from calculus but we do not need to know calculus to use them. Using calculus, areas are estimated by adding up the areas of many small rectangles to approximate the total area.



There are many continuous probability distribution used to model different situations.
We will work with three distributions in chapters 5 and 6 and will learn others later on.

- Chapter 5: Uniform Distribution and Exponential Distribution
- Chapter 6: Normal Distribution

Continuous Probability Distributions Notes, by Roberta Bloom De Anza College. This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License..
Some material is derived from Introductory Statistics from Open Stax (Illowsky/Dean) available for download for free at http://cnx.org/content 11562/latest/ or https://openstax.org/details/introductory-statistics

CHAPTER 5: UNIFORM DISTRIBUTION

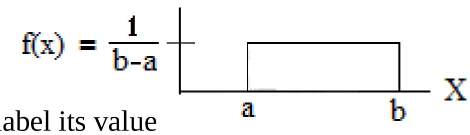
PROPERTIES OF THE UNIFORM DISTRIBUION

- continuous
- random variable can take on any value between a minimum value of a to a maximum value of b.
- All equal sized intervals of X are equally likely.
- **Notation: X ~ U(a, b)   The random variable X is distributed uniformly with between a and b.**
- Mean (average, expected value) $\mu = \dfrac{(a+b)}{2}$; Standard Deviation: $\sigma = \dfrac{(b-a)}{\sqrt{12}}$

Graph of the Uniform Distribution:

- Graph of uniform distribution is a rectangle.
- Rectangle begins at the minimum data value of **a** and ends at the maximum data value of **b**.
- The horizontal axis represents values of the random variable, X.
- The vertical axis represents values of the **probability density function**, f(x).
- f(x) = the height of the rectangle.
- For a uniform distribution, the height f(x) of the rectangle is **ALWAYS constant**.

Drawing and Labeling the Graph:
Must use a ruler to draw both axes and rectangle.
Label horizontal axis X and label the vertical axis f(x).
Mark the height of the rectangle on the vertical axis and label its value



Calculating the height of the rectangle:
The total area of the rectangle equals 1, the total probability of the variable X.
area of rectangle = base • height = 1
(b – a) • f(x) = 1
f(x) = 1/(b – a) = height of the rectangle

UNIFORM EXAMPLE 1: A delivery company divides their packages into weight classes. Suppose packages in the 14 to 20 pound class are uniformly distributed, meaning that all weights within that class are equally likely to occur. We're interested in studying packages with weights in the 14 to 20 pound class.

a. Define the random variable in words

X = ____

b. Write the distribution of X.   X ~ ______________________________

c. Neatly sketch a fully labeled graph of X. Use a ruler!

d. Find the probability that the package weighs BETWEEN 15 and 16.5 pounds.
Draw, shade, and label the graph and show your work.
Write your answer as a mathematical probability statement using correct notation.

UNIFORM EXAMPLE 1 continued:

For parts e, f, g, h Draw, shade, and label the graph and show your work.
Write your answer as a mathematical probability statement using correct notation.

e. Find the probability that a package weighs AT MOST 15 pounds.

f. Find the probability that a package weighs AT LEAST 18 pounds.

g. Find the probability that a package weighs EXACTLY 17 pounds.

h. Inverse Probability: Percentiles: Find the 40th percentile of weights of packages in this weight class.

- **Recall the interpretation of percentiles we learned from descriptive statistics**
  p% of data values are less than (or equal to) the pth percentile
  (100-p)% of data values are greater than (or equal to) the pth percentile
- The **pth percentile** is the x value dividing data between the lower p% and the upper (100 – p)% of data.
- Percentiles represent area to the left in the graph of a continuous probability distribution. pth percentile tells us that p/100 is the size of the "area to the left" on the graph of a continuous probability distribution.
  We need to find the x value at the right boundary of that area.

i. Find the mean and standard deviation to 4 decimal places for X.

Mean:

Standard Deviation:

UNIFORM EXAMPLE 2: OPTIONAL PRACTICE

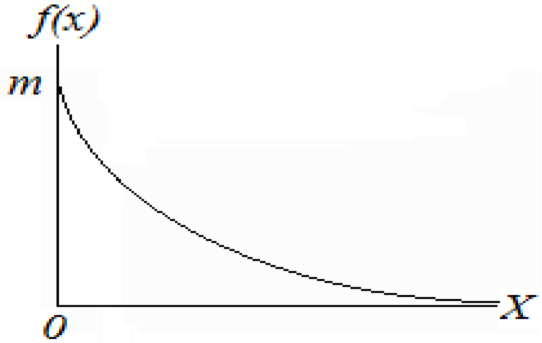
The time between busses on Stevens Creek Blvd is 12 minutes. Therefore the wait time of a passenger who arrives randomly at a bus stop is uniformly distributed between 0 and 12 minutes.

a. Find the probability that a person randomly arriving at the bus stop to wait for the bus has a wait time of at most 5 minutes.

b. Find the 80^{th} percentile of wait times for this bus, for people who arrive randomly at the bus stop.

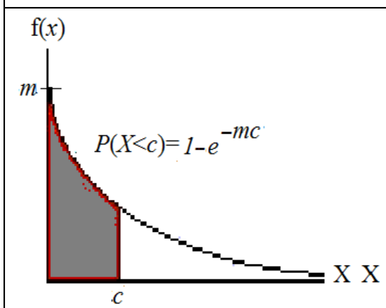
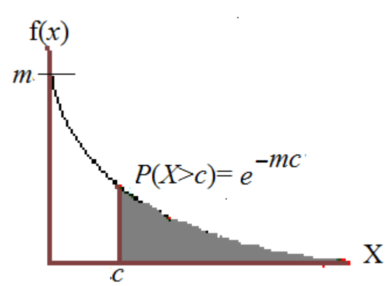
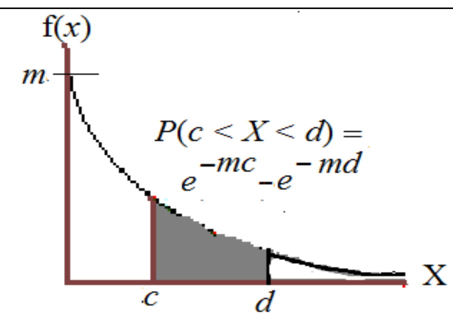
c. Find the mean and standard deviation.

CHAPTER 5: EXPONENTIAL DISTRIBUTION



- continuous function
- decreasing function
- skewed to the right
- domain: $X \ge 0$
- x-axis is a horizontal asymptote;
  the exponential curve stays above the x-axis and never touches the x axis, but gets closer and closer to the x axis as x gets large.

- **Notation: $\mathbf{X \sim Exp}\ (\boldsymbol{m})$** "random variable X is distributed exponentially with parameter *m*"
- ***m*** **is called DECAY PARAMETER OF** or **DECAY RATE**.
- m and μ are reciprocals: mean $\mu = \dfrac{1}{m}$, and $m = \dfrac{1}{\mu}$; standard deviation: $\sigma = \mu = \dfrac{1}{m}$
- Total area under the entire curve is 1 even though the area has an infinite right tail.
- Value of the Vertical Intercept is ***m***.
- Probability density function (PDF) draws the curve : $f(x) = me^{-mx}$, $m > 0$, and $x \ge 0$; $e \approx 2.71828$.
- Cumulative distribution function (CDF) (from calculus) gives 3 formulas to use to find area = probability

<table>
<tr><th colspan="3">FORMULAS for finding AREAS under the exponential curve</th></tr>
<tr><td>Area to the left of c:</td><td>Area to the right of c:</td><td>Area between two values of x, c and d:</td></tr>
<tr><td>$P(X < c) = 1 - e^{-mc}$
and
$P(X \le c) = 1 - e^{-mc}$</td><td>$P(X > c) = e^{-mc}$
and
$P(X \ge c) = e^{-mc}$</td><td>$P(c < X < d) = e^{-mc} - e^{-md}$
and
$P(c \le X \le d) = e^{-mc} - e^{-md}$</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
</table>

The exponential distribution models waiting time between random events, such as

- time until the next earthquake
- time between calls to an emergency response number
- time between calls to a customer service number
- time until failure (useful lifetime) for certain types of electrical components or other items that generally fail due to breakage but not due to wearing out; it is used by quality control engineers in high-tech hardware companies.

CHAPTER 5: EXPONENTIAL DISTRIBUTION

For each of the following examples, find the distribution and find m, μ, and σ.
Draw shade and label the graph and do the calculations to find the requested probability.
Write your answer as a mathematical probability statement.

EXPONENTIAL EXAMPLE 1: Suppose that the time between earthquakes of magnitude 5 or higher in a certain region follows an exponential distribution with an average of 40 years.

X = time between earthquakes of magnitude 5 or higher

a. X~_______________ *m*= _________ μ=_____________ σ=_____________

How did we find *m* and μ in this problem?

For parts b, c, d: Draw, shade, and label the graph and show your work.
Write your answer as a mathematical probability statement using correct notation.
State probabilities in decimal form rounded to 3 decimal places.

b. Find the probability that the time between earthquakes magnitude 5 or higher is at most 20 years.

c. Find the probability that the time between earthquakes magnitude 5 or higher is more than 25 years.

d. Find the probability that the time between earthquakes magnitude 5 or higher is between 10 and 25 years.

CHAPTER 5: EXPONENTIAL DISTRIBUTION

EXPONENTIAL EXAMPLE 2: Suppose that the time, in months, that an certain type of electronic component lasts, until if fails, is exponentially distributed with a decay parameter (or decay rate) of 0.03125.

X = time until failure for the computer component

a. X~_______________ *m*= _________ μ=_____________ σ =_____________

How did we find *m* and μ in this problem?

b. This type of electrical component lasts_____ months, on average.

For part c: Draw, shade, and label the graph and show your work.
Write your answer as a mathematical probability statement using correct notation.
State probabilities in decimal form rounded to 3 decimal places.

c. Find the probability that the time until failure for this component is less than 1 year (12 months).

***EXPONENTIAL EXAMPLE 3 OPTIONAL PRACTICE*:** A large city does a study of "911" calls to their emergency response call center. They find that the time between calls follows an exponential distribution with a mean of 5 minutes.

X = ____

a. X~_______________ *m*= _________ μ=_____________ σ =_____________

For parts b and c: Draw, shade, and label the graph and show your work.
Write your answer as a mathematical probability statement using correct notation.
State probabilities in decimal form rounded to 3 decimal places.

b. Find the probability that the time until the next "911" call is at least 10 minutes

c. Find the probability that the time until the next "911" call is between 4 and 8 minutes.

d. Find the probability that the time until the next "911" call is less than 5 minutes

CHAPTER 5: INVERSE PROBABILITY for EXPONENTIAL DISTRIBUTION

- **Recall the interpretation of percentiles we learned from descriptive statistics**
  p% of data values are less than (or equal to) the pth percentile
  (100-p)% of data values are greater than (or equal to) the pth percentile
- The **p^{th} percentile** is the x value dividing data between the lower p% and the upper (100 – p)% of data.
- Percentiles represent area to the left in the graph of a continuous probability distribution. p^{th} percentile tells us that p/100 is the size of the "area to the left" on the graph of a continuous probability distribution.
  We need to find the x value at the right boundary of that area.

To find the percentile we use the formula for probability for area to the left and solve for *k*.
Because *k* is in the exponent, we need to use natural log, ln.

$P(X < k) = 1 - e^{-mk}$

Area to the left $= 1 - e^{-mk}$

Formula for Inverse Probability for Exponential Distribution

If $P(X < k)$ = area to the left, then $k = \dfrac{\ln(1 - \textit{area to the left})}{(-m)}$

OPTIONAL: EXPONENTIAL EXAMPLE 4: Inverse: A large city does a study of "911" calls to their emergency response call center. They find that the time between calls follows an exponential distribution with a mean of 5 minutes. Find the 60^{th} percentile of times between "911" calls.

OPTIONAL: EXPONENTIAL EXAMPLE 5: Inverse: Suppose X~Exp (0.1)

a. Draw the graph that represents the median and find the value of the median.

b. Find the mean. Does the mean equal the median? Explain why or why not based on the shape of the graph.