Chapter 4.5, 6, 8 Probability Distributions for Continuous Random Variables

- Discrete vs. continuous random variables
- Examples of continuous distributions
  - Uniform
  - Exponential
  - Normal

Recall:
A random variable = numerical description of outcomes of a random experiment. Assigns a number to each outcome.
- **Discrete** - one can count and list the possible values
- **Continuous** - possible values are all real numbers in an interval

The graphical form of the probability distribution for a discrete random variable \( x \) is a histogram. If relative frequencies are used, the area of all bars is 1 (100% of a unity)

The areas in the bars represent probabilities for \( x \) from the classes represented by the bars.

The graphical form of the probability distribution for a continuous random variable \( x \) can be represented by a smooth curve. This curve is called a Probability Density Function (PDF)
The area between the curve and the horizontal axis is 1 square unit (100% of a unity)

The areas under a probability distribution curve correspond to probabilities for \( x \) from given interval of data: \( P(a < x < b) \) is the area under the curve for the interval from \( a \) to \( b \).
Chapter 4.6 The Normal Distribution

Properties of the Normal Distribution

The normal is a family of

✓ Bell-shaped, symmetric distributions.
✓ Characterized by a mean, \( \mu \), and variance, \( \sigma^2 \). That is: \( X \sim N(\mu, \sigma^2) \).
✓ Each is asymptotic to the horizontal
✓ If several independent random variables are normally distributed then their sums or differences will also be normally distributed.

Pdf formula is

\[
 f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}
\]

where

\( \mu \) = Mean of the normal random variable \( x \)
\( \sigma \) = Standard deviation
\( \pi \) = 3.1415 . . .
\( e \) = 2.71828 . . .
\( P(x < a) \) is obtained from a table of normal probabilities or from software or a calculator like a TI

The values of the mean and standard deviation affect the shape and position of the graph:

The probability that \( x \) assumes the value between \( a \) and \( b \) can be computed as the area under the curve:

\[
 P(c \leq x \leq d) = \int_{c}^{d} f(x) \, dx
\]

But we WON'T compute such sophisticated integrals here! We'll use software or calculators.
The **standard normal distribution** is a normal distribution with $\mu = 0$ and $\sigma = 1$. A random variable with a standard normal distribution, denoted by the symbol $z$, is called a **standard normal random variable**.

### Finding the Probabilities for SND:

**Example:** Find the probability that random variable $z$ is between 0 and 1.56

$P(0 \leq z \leq 1.56)$

- **Tables:** Look in row labeled 1.5 and column labeled .06 to find
  
  $P(0 \leq z \leq 1.56) = 0.4406$

- **Calculator:** `Normalcdf(0, 1.56, 0, 1)`

### Standardizing Normal Distribution:

If $x$ is a normal random variable with mean $\mu$ and standard deviation $\sigma$, then the random variable $z$, defined by the formula

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

has a standard normal distribution. The value $z$ describes the number of standard deviations between $x$ and $\mu$.

**Example: Using standard normal distribution to compute probabilities**

$X \sim N(160, 30^2)$. Find $P(100 < x < 180)$

$$P(100 \leq X \leq 180) = P\left(\frac{100 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{180 - \mu}{\sigma}\right)$$

$$= P\left(\frac{100 - 160}{30} \leq Z \leq \frac{180 - 160}{30}\right)$$

$$= P\left(-2 \leq Z \leq 0.666\right)$$

$$= 0.477 + 0.247 = 0.724$$

(see the tables)

Simpler: use the calculator

`Normalcdf(100, 180, 160, 30)`
More examples:
\[ P(z<-2.5) = \text{normalcdf}(-10^9, -2.5, 0, 1) \]
\[ P(z>1.96) = \text{normalcdf}(1.96, 10^9, 0, 1) \]
\[ P(0.50<z<1.50) = \text{normalcdf}(0.50, 1.50, 0, 1) \]

Exercises:
Exercise 1. Let \( z \) be a standard normal random variable.
1. Find the following probabilities using TI-83. Draw the area that represents the probability.
   a. \( P(z > -1.25) \)
   b. \( P(z \leq -0.63) \)
   c. \( P(-1.10 < z \leq 1.63) \)
   d. \( P(|z| > 1.59) \)
   e. \( P(z = 2.00) \)

Exercise 2. Let \( x \) be normal random variable with \( \mu = 30 \) and \( \sigma = 4 \).
1. Find the following probabilities using TI-83. Draw the area that represents the probability.
   a. \( P(27 < x \leq 41) \)
   b. \( P(x > 25) \)
   c. \( P(x \leq 40) \)
   d. \( P(x \leq 1.59) \)

Reversing the problem: finding the percentiles
given is the probability, find a cut-off score (a percentile)
Find the percentiles: \( P_{25}, P_{50}, P_{90} \) (tables and/or a calculator)
TI-83: \text{InverseNorm}(0.25, 0, 1)\) Answer: \( P_{25} = -0.674 \) (can be rounded to the hundredths)
Interpretation:

How to – TI:
\text{InverseNorm(area in the left corner in decimal, the mean, standard deviation)}

Example: Find the scores separating 95% of most frequent scores of a random variable modeled by the standard normal distribution
Solution:

**Exercises 4.84–4.116**

**Learning the Mechanics**

4.84 Find the area under the standard normal probability distribution between the following pairs of z-scores:
   a. $z = 0$ and $z = 2.00$
   b. $z = 0$ and $z = 3$

4.86 Find the following probabilities for the standard normal random variable $z$:
   a. $P(-1 < z < 1)$
   e. $P(z \geq -2.33)$

4.87 Find each of the following probabilities for the standard normal random variable $z$:
   a. $P(-1 \leq z \leq 1)$
   b. $P(-1.96 \leq z \leq 1.96)$
   c. $P(-1.645 \leq z \leq 1.645)$
   d. $P(-2 \leq z \leq 2)$

Illustrate the number on the graph of a normal curve.

4.88 Find a value of the standard normal random variable $z$, call it $z_0$, such that
   a. $P(z \geq z_0) = 0.05$
   b. $P(z \geq z_0) = 0.025$
   c. $P(z \leq z_0) = 0.025$

4.89 Find a value of the standard normal random variable $z$, call it $z_0$, such that
   a. $P(z \leq z_0) = 0.2090$
   c. $P(-z_0 \leq z < z_0) = 0.8472$
   e. $P(z \leq z_0 \leq z \leq 0) = 0.4798$

More: Find a value $z_0$ such that the condition below is met.
   a. $P(z \leq z_0) = 0.15$
   b. $P(z > z_0) = 0.79$
   c. $P(|z| \leq z_0) = 0.60$
4.100 The business of casino gaming. Casino gaming yields over $35 billion in revenue each year in the United States. In *Chance* (Spring 2005), University of Denver statistician R. C. Hannum discussed the business of casino gaming and its reliance on the laws of probability. Casino games of pure chance (e.g., craps, roulette, baccarat, and keno) always yield a “house advantage.” For example, in the game of double-zero roulette, the expected casino win percentage is 5.26% on bets made on whether the outcome will be either black or red. (This implies that for every $5 bet on black or red, the casino will earn a net of about 25¢.) It can be shown that in 100 roulette plays on black/red, the average casino win percentage is normally distributed with mean 5.26% and standard deviation 10%. Let $x$ represent the average casino win percentage after 100 bets on black/red in double-zero roulette.

**a.** Find $P(x>0)$. (This is the probability that the casino wins money.)

**b.** Find $P(5<x<15)$.

**c.** Find $P(x<1)$.

**d.** If you observed an average casino win percentage of −25% after 100 roulette bets on black/red, what would you conclude?

Approximating Binomial Distribution by Normal

Binomial distribution of $x$ successes in a sample size $n$ with probability of success $p$ has the mean $\mu=np$ and the variance $\sigma^2=np(1-p)$.

For large $n$ the binomial distribution can be approximated by normal distribution with the same mean and variance (more precise conditions when we can use this approximation will be given in the next chapter).

4.104 LASIK surgery complications. According to studies, 1% of all patients who undergo laser surgery (i.e., LASIK) to correct their vision have serious postlaser vision problems (*All About Vision*, 2012). In a sample of 100,000 patients, what is the approximate probability that fewer than 950 will experience serious postlaser vision problems?

Chapter 4.8 Other Continuous Distributions
**Uniform distribution**

All outcomes have the same chance to occur.

The graph of the uniform distribution is a horizontal line segment. It defines the area between itself and the horizontal axis.

The function modeling the distribution is called the

Probability Density Function (pdf):

\[ f(x) = \frac{1}{d-c} \quad c \leq x \leq d \]

Mean: \[ \mu = \frac{c + d}{2} \]

Standard Deviation: \[ \sigma = \frac{d - c}{\sqrt{12}} \]

**Example:** Suppose the time that you wait on the telephone for a live representative of your phone company to discuss your problem is uniformly distributed between 5 and 25 minutes.

a. Plot a graph for this distribution.

b. What is the probability that you’ll wait less than 6 minutes?

c. What is the probability that you’ll wait between 10 and 15 minutes?

d. How many minutes of waiting separate 10% of the longest waiting time?

e. What is the mean and standard deviation of this distribution?

For each problem re-draw the graph of this distribution and shade an appropriate part of the graph or mark and label the answer on data line.
**Exponential Distribution.** Exponential distribution is used to model the distribution of the length of time (or distance, or area, ...) between two consecutive (rare) events in Poisson experiment.

![Exponential Distribution Diagram](image)

Examples of random variables with exponential distributions:
- The length of time between emergency arrivals at a hospital
- The length of time between breakdowns of manufacturing equipment
- The length of time between catastrophic events (e.g., a stock market crash)

This distribution is called the **waiting time distribution**.

**Probability Distribution for an Exponential Random Variable** $x$

PDF: \[ f(x) = \frac{1}{\theta} e^{-x/\theta} \quad (x > 0) \]

Mean $\mu = \theta$, Standard Deviation $\sigma = \theta$

**Example.** Suppose the length of time (in hours) between emergency arrivals at a certain hospital is modeled as an exponential distribution with $\theta = 2$. What is the probability that more than 5 hours pass without an emergency arrival?

Solution: $f(x) = 0.5e^{-x/2}$, $x > 0$, is a pdf (probability density function) of an exponential random variable with $\mu = \theta = 2$

Computation technique employs the following fact from *Calculus*

\[ A = P(x \geq a) = e^{-a/\theta} \]
\[ A = e^{-\frac{a}{\theta}} = e^{-\left(\frac{5}{2}\right)} = e^{-2.5} \]

There is about 8.2% chance that more than 5 hours will pass between emergency arrivals.

**Exercises.** Suppose that \( x \) has an exponential distribution with mean \( \mu = 3 \). Then \( \theta = \mu = 3 \),

\[ P(x > 2) = P(x \geq 2) = e^{2/3} = .5134 \]

\[ P(x < 5) = 1 - P(x \geq 5) = 1 - e^{5/3} = .8111 \]

**4.140 Preventative maintenance tests.** The optimal scheduling of preventative maintenance tests of some (but not all) of \( n \) independently operating components was developed in *Reliability Engineering and System Safety* (Jan. 2006). The time (in hours) between failures of a component was approximated by an exponential distribution with mean \( \theta \).

a. Suppose \( \theta=1,000 \) hours. Find the probability that the time between component failures ranges between 1,200 and 1,500 hours.

b. Again, assume \( \theta=1,000 \) hours. Find the probability that the time between component failures is at least 1,200 hours.

c. Given that the time between failures is at least 1,200 hours, what is the probability that the time between failures is less than 1,500 hours?

**Brief Summary Chapter 4**

**Discrete Distributions:**
- Binomial
- Hypergeometric
- Poisson

**Continuous Distributions:**
- Normal
- Uniform
- Exponential