MSSC 6010 / Comp. Probability

Instructor: Mehdi Maadooliat

Chapter 3

Department of Mathematics, Statistics and Computer Science

Special thanks to Prof. Ana Militino for providing the original slides of the book.

Chapter 3

General Probability and Random Variables

3.3 Probability

3.3.1 Sample Space and Events

- An **experiment** is any action or process that generates observations.
- The **sample space** of an experiment, denoted by Ω, is the set of all of the possible outcomes of an experiment.

1

1

- Although the outcome of an experiment cannot be known before it has taken place, it is possible to define the sample space for a given experiment. The sample space may be either finite or infinite.
- For example, the number of unoccupied seats in a train corresponds to a finite sample space. The number of passengers arriving at an airport also produces a finite sample space, assuming a one to one correspondence between arriving passengers and the natural numbers.
- The sample space for the lifetime of light bulbs, however, is infinite, since lifetime may be any positive value.
- An **event** is any subset of the sample space, which is often denoted with the letter E.

- Events are said to be **simple** when they contain only one outcome; otherwise, events are considered to be **compound**. Consider an experiment where a single die is thrown.
- Since the die might show any one of six numbers, the sample space is written $\Omega = \{1, 2, 3, 4, 5, 6\}$; and any subset of Ω, such as $E_1 = \{\text{even numbers}\}$, $E_2 = \{2\}$, $E_3 = \{1, 2, 4\}$, $E_4 = \Omega$, or $E_5 = \emptyset$, is considered an event.
- Specifically, E_2 is considered a simple event while all of the remaining events are considered to be compound events.
- Event E_5 is known as the **empty set** or the **null set**, the event that does not contain any outcomes.
- In many problems, the events of interest will be formed through a combination of two or more events by taking **unions**, **intersections**, and **complements**.

3.3.2 Set Theory

- The following definitions review some basic notions from set theory and some basic rules of probability that are not unlike the rules of algebra.
- For any two events E and F of a sample space Ω, define the new event $E \cup F$ (read E union F) to consist of all outcomes that are either in E or in F or in both E and F.
- In other words, the event $E \cup F$ will occur if either E or F occurs.
- In a similar fashion, for any two events E and F of a sample space Ω, define the new event $E \cap F$ (read E intersection F) to consist of all outcomes that are both in E and in F.
- Finally, the complement of an event E (written E^c) consists of all outcomes in Ω that are not contained in E.

- Given events $E, F, G, E_1, E_2, \ldots,$ the commutative, associative, distributive, and DeMorgan's laws work as follows with the union and intersection operators:

1. Commutative laws
   - for the union $\quad E \cup F = F \cup E$
   - for the intersection $\quad E \cap F = F \cap E$
2. Associative laws
   - for the union $\quad (E \cup F) \cup G = E \cup (F \cup G)$
   - for the intersection $\quad (E \cap F) \cap G = E \cap (F \cap G)$
3. Distributive laws
   - $(E \cap F) \cup G = (E \cup G) \cap (F \cup G)$
   - $(E \cup F) \cap G = (E \cap G) \cup (F \cap G)$
4. Demorgan's laws
   - $\left(\bigcup_{i=1}^{\infty} E_i\right)^c = \bigcap_{i=1}^{\infty} E_i^c$
   - $\left(\bigcap_{i=1}^{\infty} E_i\right)^c = \bigcup_{i=1}^{\infty} E_i^c$

3.3.3 Interpreting Probability

3.3.3.1 Relative Frequency Approach to Probability

Suppose an experiment can be performed n times under the same conditions with sample space, Ω. Let $n(E)$ denote the number of times (in n experiments) that the event E occurs. The relative frequency approach to probability defines the probability of the event E, written $\mathbb{P}(E)$, as

$$\mathbb{P}(E) = \lim_{n\to\infty} \frac{n(E)}{n}.$$

Although the preceding definition of probability is intuitively appealing, it has a serious drawback. There is nothing in the definition to guarantee $\frac{n(E)}{n}$ converges to a single value.

3.3.3.2 Axiomatic Approach to Probability

The Three Axioms of Probability

Consider an experiment with sample space, Ω. For each event E of the sample space Ω, assume that a number $\mathbb{P}(E)$ is defined that satisfies the following three axioms:

1. $0 \le \mathbb{P}(E) \le 1$
2. $\mathbb{P}(\Omega) = 1$
3. For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is $E_i \cap E_j = \emptyset$) for all $i \neq j$,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i).$$

The following results are all easily derived using some combination of the three axioms of probability.

1. $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$
   Proof: Note that E and E^c are always mutually exclusive. Since $E \cup E^c = \Omega$, by axiom 2 on the preceding page and axiom 3 on the prev
   $1 = \mathbb{P}(\Omega) = \mathbb{P}(E \cup E^c) = \mathbb{P}(E) + \mathbb{P}(E^c)$.

2. $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$

3. $\mathbb{P}(\emptyset) = 0$

4. If $E \subset F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$

Example 3.9 ▷ ***Law of Complement: Birthday Problem*** ◁ Suppose that a room contains m students. What is the probability that at least two of them have the same birthday? This is a famous problem with a counterintuitive answer. Assume that every day of the year is equally likely to be a birthday, and disregard leap years. That is, assume there are always $n = 365$ days to a year.

Solution: Let the event E denote two or more students with the same birthday. In this problem, it is easier to find E^c, as there are a number of ways that E can take place. There are a total of 365^m possible outcomes in the sample space. E^c can occur in $365 \times 364 \times \cdots \times (365 - m + 1)$ ways. Consequently,

$$\mathbb{P}(E^c) = \frac{365 \times 364 \times \cdots \times (365 - m + 1)}{365^m}$$

and

$$\mathbb{P}(E) = 1 - \frac{365 \times 364 \times \cdots \times (365 - m + 1)}{365^m}.$$

The following S code can be used to create or modify a table such as the one in Table 3.1 on the facing page, which gives $\mathbb{P}(E)$ for $m = 10, 15, \ldots, 50$.

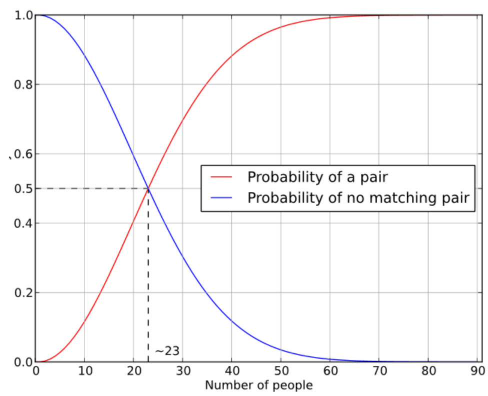
```
➢ for (m in seq(10,50,5))
      print(c(m, 1 - prod(365:(365-m+1))/365^m))
```

Another approach that can be used to solve the problem is to enter

```
➢ m <- seq(10,50,5)
➢ P.E <- function(m){c(m,1-prod(365:(365-m+1))/365^m)}
➢ t(sapply(m,P.E))
```


BIRTHDAY PARADOX

- **What's the chances that two people in our class have the same birthday?**



3.3.4 Conditional Probability If E and F are any two events in a sample space Ω and $\mathbb{P}(E) \neq 0$, the **conditional probability** of F given E is defined as

$$\mathbb{P}(F|E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)}. \tag{3.1}$$

It is left as an exercise for the reader to verify that $\mathbb{P}(F|E)$ satisfies the three axioms of probability.

Example 3.11 Suppose two fair dice are tossed where each of the 36 possible outcomes is equally likely to occur. Knowing that the first die shows a 4, what is the probability that the sum of the two dice equals 8?

Solution: The sample space for this experiment is given as $\Omega = \{(i, j),\ i = 1, 2, \ldots, 6,\ \ j = 1, 2, \ldots, 6\}$ where each pair (i, j) has a probability $1/36$ of occurring. Define "the sum of the dice equals 8" to be event F and "a 4 on the first toss" to be event E. Since $E \cap F$ corresponds to the outcome $(4, 4)$ with probability $\mathbb{P}(E \cap F) = 1/36$ and there are six outcomes with a 4 on the first toss, $(4, 1), (4, 2), \ldots, (4, 6)$, the probability of event E, $\mathbb{P}(E) = 6/36 = 1/6$ and the answer is calculated as

$$\mathbb{P}(F|E) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(E)} = \frac{1/36}{1/6} = \frac{1}{6}. \qquad \blacksquare$$

Example 3.12 Suppose a box contains 50 defective light bulbs, 100 partially defective light bulbs (last only 3 hours), and 250 good light bulbs. If one of the bulbs from the box is used and it does not immediately go out, what is the probability the light bulb is actually a good light bulb?

Solution: The conditional probability the light bulb is good given that the light bulb is not defective is desired. Using (3.1), write:

$$\mathbb{P}(\text{Good}|\text{Not Defective}) = \frac{\mathbb{P}(\text{Good})}{\mathbb{P}(\text{Not Defective})} = \frac{250/400}{350/400} = \frac{5}{7}. \quad \blacksquare$$

3.3.5 The Law of Total Probability and Bayes' Rule

Law of Total Probability — Let $F_1, F_2, \ldots, F_n$ be such that $\bigcup_{i=1}^{n} F_i = \Omega$ and $F_i \cap F_j = \emptyset$ for all $i \neq j$, with $\mathbb{P}(F_i) > 0$ for all i. Then, for any event E,

$$\mathbb{P}(E) = \sum_{i=1}^{n} \mathbb{P}(E \cap F_i) = \sum_{i=1}^{n} \mathbb{P}(E|F_i)\mathbb{P}(F_i). \tag{3.2}$$

At times, it is much easier to calculate the conditional probabilities $\mathbb{P}(E|F_i)$ for an appropriately selected F_i than it is to compute $\mathbb{P}(E)$ directly. When this happens, **Bayes' Rule** is used, which is derived using (3.1), to find the answer.

Bayes' Rule — Let $F_1, F_2, \ldots, F_n$ be such that $\bigcup_{i=1}^{n} F_i = \Omega$ and $F_i \cap F_j = \emptyset$ for all $i \neq j$, with $\mathbb{P}(F_i) > 0$ for all i. Then,

$$\mathbb{P}(F_j|E) = \frac{\mathbb{P}(E \cap F_j)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E|F_j)\mathbb{P}(F_j)}{\sum_{i=1}^{n} \mathbb{P}(E|F_i)\mathbb{P}(F_i)}. \tag{3.3}$$

Example 3.13 ▷ ***Conditional Probability: Car Batteries*** ◁ A car manufacturer purchases car batteries from two different suppliers. Supplier A provides 55% of the batteries and supplier B provides the rest. If 5% of all batteries from supplier A are defective and 4% of the batteries from supplier B are defective, determine the probability that a randomly selected battery is not defective.

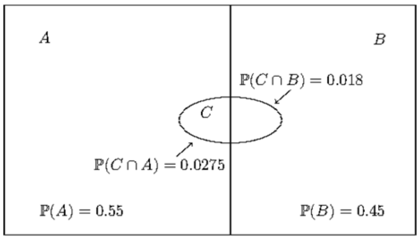


Figure 3.1: Sample space for Example 3.13

Solution: Let C correspond to the event "the battery does not work properly," A to the event "the battery was supplied by A," and B to the event "the battery was supplied by B." Since a working battery might come from either supplier A or B, A and B are disjoint events.

Consequently, $\mathbb{P}(C) = \mathbb{P}(C \cap A) + \mathbb{P}(C \cap B)$. Given that

$$\begin{array}{ll} \mathbb{P}(A) = 0.55,\ \mathbb{P}(C|A) = 0.05, & \mathbb{P}(C \cap A) = \mathbb{P}(C|A)\mathbb{P}(A), \\ \mathbb{P}(B) = 0.45,\ \mathbb{P}(C|B) = 0.04,\ \text{and} & \mathbb{P}(C \cap B) = \mathbb{P}(C|B)\mathbb{P}(B), \end{array}$$

write $\mathbb{P}(C) = (0.05)(0.55)+(0.04)(0.45) = 0.0455$. Then, the probability that the battery works properly is $1 - \mathbb{P}(C) = 0.9545$. ■

Example 3.14 Suppose a student answers all of the questions on a multiple-choice test. Let p be the probability the student actually knows the answer and $1-p$ be the probability the student is guessing for a given question. Assume students that guess have a $1/a$ probability of getting the correct answer, where a represents the number of possible responses to the question. What is the conditional probability a student knew the answer to a question given that he answered correctly?

Solution: Let the event E, "question answered correctly," F_1, represent the events "student knew the correct answer," and F_2 "student guessed" respectively. Using (3.3) write

$$\mathbb{P}(F_1|E) = \frac{\mathbb{P}(F_1 \cap E)}{\mathbb{P}(E)} = \frac{\mathbb{P}(F_1)}{\mathbb{P}(E|F_1)\mathbb{P}(F_1) + \mathbb{P}(E|F_2)\mathbb{P}(F_2)}$$
$$= \frac{\mathbb{P}(F_1)}{\mathbb{P}(F_1) + \mathbb{P}(F_2 \cap E)}$$
$$= \frac{p}{p + (1-p)/a}$$

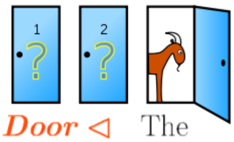
because $\mathbb{P}(E|F_2) = \dfrac{1}{a}$, and $\mathbb{P}(F_2) = 1 - p$

$\mathbb{P}(E|F_2) = \dfrac{\mathbb{P}(F_2 \cap E)}{\mathbb{P}(F_2)}$ then $\mathbb{P}(F_2 \cap E) = \dfrac{1-p}{a}$

As a special case, if $a = 4$ and $p = 1/2$, then the probability a student actually knew the answer given their response was correct is $4/5$. ■

LET'S MAKE A DEAL
(MONTY HALL PROBLEM)

- http://en.wikipedia.org/wiki/Monty_Hall_problem



Example 3.15 ▷ ***Bayes' Rule: Choose a Door*** ◁ The television show *Let's Make a Deal* hosted by Monty Hall gave contestants the chance to choose, among three doors, the one that concealed the grand prize. Behind the other two doors were much less valuable prizes. After the contestant chose one of the doors, say Door 1, Monty opened one of the other two doors, say Door 3, containing a much less valuable prize. The contestant was then asked whether he or she wished to stay with the original choice (Door 1) or switch to the other closed door (Door 2). What should the contestant do? Is it better to stay with the original choice or to switch to the other closed door? Or does it really matter? The answer, of course, depends on whether contestants improve their chances of winning by switching doors. In particular, what is the probability of winning by switching doors when given the opportunity; and what is the probability of

20

winning by staying with the initial door selection? First, simulate the problem with S to provide approximate probabilities for the various strategies. Following the simulation, show how Bayes' Rule can be used to solve the problem exactly.

Solution: To simulate the problem, generate a random vector named `actual` of size 10,000 containing the numbers 1, 2, and 3. In the vector `actual`, the numbers 1, 2, and 3 represent the door behind which the grand prize is contained. Then, generate another vector named `guess` of size 10,000 containing the numbers 1, 2, and 3 to represent the contestant's initial guess. If the i^{th} values of the vectors `actual` and `guess` agree, the contestant wins the grand prize by staying with his initial guess. On the other hand, if the i^{th} values of the vectors `actual` and `guess` disagree, the contestant wins the grand prize by switching. Consider the following S code and the

21

results which suggest the contestant is twice as likely to win the grand prize by switching doors.

```
➢ actual <- sample(1:3, 10000, replace = T)
➢ aguess <- sample(1:3, 10000, replace = T)
➢ equals <- (actual == aguess)
➢ PNoSwitch <- sum(equals)/10000
➢ not.eq <- (actual != aguess)
➢ PSwitch <- sum(not.eq)/10000
➢ Probs <- c(PNoSwitch, PSwitch)
➢ names(Probs) <- c("P(Win no Switch)", "P(Win
  Switch)")
➢ Probs
      P(Win no Switch) P(Win Switch)
            0.3317        0.6683
```


Next use (3.3) after defining events D_i and O_j to find $\mathbb{P}(D_1|O_3)$ and $\mathbb{P}(D_2|O_3)$. Start by assuming the contestant initially guesses door 1 and that Monty opens door 3.

Let the event D_i = door i conceals the prize and O_j = Monty opens door j after the contestant selects door 1.

When a contestant initially selects a door, $\mathbb{P}(D_1) = \mathbb{P}(D_2) = \mathbb{P}(D_3) = 1/3$.

Once Monty shows the grand prize is not behind door 3, the probability of winning the grand prize is now one of $\mathbb{P}(D_1|O_3)$ or $\mathbb{P}(D_2|O_3)$.

Note that $\mathbb{P}(D_1|O_3)$ corresponds to the strategy of sticking with the initial guess and $\mathbb{P}(D_2|O_3)$ corresponds to the strategy of switching doors. Based on how the show is designed the following are known:

- $\mathbb{P}(O_3|D_1) = 1/2$ since Monty can open one of either door 3 or door 2.
- $\mathbb{P}(O_3|D_2) = 1$ since the only door Monty can open without revealing the grand prize is door 3.
- $\mathbb{P}(O_3|D_3) = 0$ since Monty will not open door 3 if it contains the grand prize.

$$\begin{aligned}\mathbb{P}(D_1|O_3) &= \frac{\mathbb{P}(O_3|D_1)\mathbb{P}(D_1)}{\mathbb{P}(O_3|D_1)\mathbb{P}(D_1) + \mathbb{P}(O_3|D_2)\mathbb{P}(D_2) + \mathbb{P}(O_3|D_3)\mathbb{P}(D_3)} \\ &= \frac{1/2 \times 1/3}{1/2 \times 1/3 + 1 \times 1/3 + 0 \times 1/3} = \frac{1}{3}\end{aligned}$$

$$\begin{aligned}\mathbb{P}(D_2|O_3) &= \frac{\mathbb{P}(O_3|D_2)\mathbb{P}(D_2)}{\mathbb{P}(O_3|D_1)\mathbb{P}(D_1) + \mathbb{P}(O_3|D_2)\mathbb{P}(D_2) + \mathbb{P}(O_3|D_3)\mathbb{P}(D_3)} \\ &= \frac{1 \times 1/3}{1/2 \times 1/3 + 1 \times 1/3 + 0 \times 1/3} = \frac{2}{3}\end{aligned}$$

Therefore, it is always to the contestant's benefit to switch doors. ∎

3.3.6 Independent Events

- Two events E and F are independent if and only if $\mathbb{P}(E|F) = \mathbb{P}(E)$ or $\mathbb{P}(F|E) = \mathbb{P}(F)$.
- An equivalent way to define independence between two events is to use (3.1) and to show that $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$.
- Independence between two events is really a special case of independence among n events. Define events $E_1, \ldots, E_n$ to be independent if for every k where $k = 2, \ldots, n$ and every subset of indices $i_1, i_2, \ldots, i_k$, $\mathbb{P}(E_{i_1} \cap E_{i_2} \cap \cdots \cap E_{i_k}) = \mathbb{P}(E_{i_1})\mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_k})$.
- It is important to point out that events in any subset of the original independent events of size r, where $r \leq k$, are also independent.
- Further, if events $E_1, \ldots, E_n$ are independent, then so are $E_1^c, \ldots, E_n^c$.

MARQUETTE UNIVERSITY
Be The Difference.

3.4 Random Variables

In general, a **random variable** is a function from a sample space Ω into the real numbers. Random variables will always be denoted with uppercase letters, for example X or Y, and the realized values of the random variable will be denoted with lowercase letters, for example x or y. Here are some examples of random variables:

1. Toss two dice. X = sum of the numbers on the dice.
2. Surgeon performs twenty heart transplants. X = number of successful transplants.
3. Individual 40 kilometer cycling time trial. X = time to complete the course.

- Random variables may be either **discrete** or **continuous**.
- A random variable is said to be discrete if its set of possible outcomes is finite or at most countable.
- If the random variable can take on a continuum of values, it is continuous.
- If Y is a random variable that is distributed approximately $DIST$ with parameter(s) $\boldsymbol{\theta}$, write $Y \stackrel{\cdot}{\sim} DIST(\boldsymbol{\theta})$.

26

MARQUETTE UNIVERSITY
Be The Difference.

47

3.4.1 Discrete Random Variables

- A discrete random variable assumes each of its values with a certain probability. When two dice are tossed, the probability the sum of two dice is seven, written $\mathbb{P}(X = 7)$, equals $1/6$.
- The function that assigns probability to the values of the random variable is called the probability density function, (**pdf**).
- Many authors also refer to the **pdf** as the probability mass function (**pmf**) when working with discrete random variables. Denote the **pdf** as $p(x) = \mathbb{P}(X = x)$ for each x.
- All **pdf**s must satisfy the following two conditions:
  1. $p(x) \geq 0$ for all x.
  2. $\sum_{\forall x} p(x) = 1$.

27

MARQUETTE UNIVERSITY
Be The Difference.

The cumulative distribution function, (**cdf**), is defined as

$$F(x) = \mathbb{P}(X \leq x) = \sum_{k \leq x} p(k).$$

Discrete **cdf**s have the following properties:

1. $0 \leq F(x) \leq 1$.
2. If $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b. In other words, $F(x)$ is a nondecreasing function of x.
3. $\lim_{x \to \infty} F(x) = 1$
4. $\lim_{x \to -\infty} F(x) = 0$
5. $F(x)$ is a step function, and the height of the step at x is equal to $f(x) = \mathbb{P}(X = x)$.

MARQUETTE UNIVERSITY
Be The Difference.

Example 3.17 Toss a fair coin three times and let the random variable X represent the number of heads in the three tosses. Produce graphical representations of both the **pdf** and **cdf** for the random variable X.

Solution: The sample space for the experiment is

$$\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$$

The random variable X can take on the values 0, 1, 2, and 3 with probabilities $\frac{1}{8}$, $\frac{3}{8}$, $\frac{3}{8}$, and $\frac{1}{8}$ respectively. Define the **cdf** for X, $F(x) =$

MARQUETTE UNIVERSITY
Be The Difference.

$\mathbb{P}(X \leq x)$ as follows:

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1/8 & \text{if } 0 \leq x < 1 \\ 4/8 & \text{if } 1 \leq x < 2 \\ 7/8 & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases}$$

To produce a graph similar to Figure 3.3 on page 52 with placement of specific values along the axes for both the **pdf** and **cdf** using the function **axis()** follows.

```
➢ x <- 0:3
➢ fx <- c(1/8,3/8,3/8,1/8)
➢ Fx <- c(1/8,4/8,7/8,1) # or Fx <- cumsum(fx)
➢ par(mfrow=c(1,2),pty="s")
```


MARQUETTE UNIVERSITY
Be The Difference.

```
➢ plot(x, fx, type="h", xlab="x", ylab="P(X=x)",
      xlim=c(0,3), ylim=c(0,.4), xaxt="n", yaxt="n")
➢ axis(1,at=c(0,1,2,3),labels=c(0,1,2,3),las=1)
➢ axis(2,at=c(1/8,3/8),labels=c("1/8","3/8"),las=1)
➢ title("PDF")
➢ plot(x, Fx,type="n", xlab="x", ylab="F(x)",
      xlim=c(-1,5), ylim=c(0,1),yaxt="n")
➢ axis(2,at=c(1/8,4/8,7/8,1),labels=c("1/8","4/8",
      "7/8","1"))
➢ segments(-1,0,0,0)
➢ segments(0:4,c(Fx,1),1:5,c(Fx,1))
➢ lines(x,Fx,type="p",pch=16)
➢ segments(-1,1,5,1,lty=2)
➢ title("CDF")
```

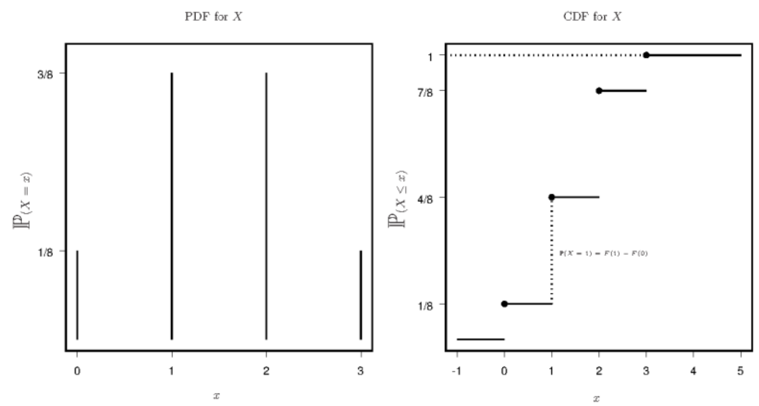


Figure 3.3: The **pdf** and **cdf** for the random variable X, the number of heads in three tosses of a fair coin

3.4.2 Mode, Median, and Percentiles

- The **mode** of a probability distribution is the x value most likely to occur. If more than one such x value exists, the distribution is multimodal.
- The **median** of a distribution is the value m such that $\mathbb{P}(X \le m) \ge 1/2$ and $\mathbb{P}(X \ge m) \ge 1/2$.
- The j^{th} **percentile** of a distribution is the value x_j such that $\mathbb{P}(X \le x_j) \ge \frac{j}{100}$ and $\mathbb{P}(X \ge x_j) \ge 1 - \frac{j}{100}$. The m value that satisfies the definition for the median is not unique.
- If Example 3.17 on page 49 is considered, the modes are 1 and 2; and any value m between 1 and 2, not inclusive, satisfies the definition for the median.
- The 25^{th} percentile of the distribution of X is 1 because $\mathbb{P}(X \le 1) = \frac{4}{8} \ge \frac{25}{100}$ and $\mathbb{P}(X \ge 1) = \frac{7}{8} \ge 1 - \frac{25}{100}$.

3.4.3 Expected Values of Discrete Random Variables

Given a discrete random variable X with **pdf** $p(x)$, the **expected value** of the random variable X, written $E[X]$, is

$$E[X] = \sum_{x} x \cdot p(x) \qquad (3.4)$$

Also denote $E[X]$ as μ_X, recognizing that $E[X]$ is the mean of the random variable X. In this definition, it is assumed the sum exists; otherwise, the expectation is undefined. It can be helpful to think of $E[X]$ as the fulcrum on a balance beam as illustrated in Figure 3.4 on the following page.

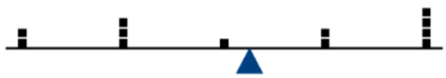


Figure 3.4: Fulcrum illustration of $E[X]$

Example 3.18 A particular game is played where the contestant spins a wheel which can lands on the numbers 1, 5, or 30 with probabilities of 0.50, 0.45, and 0.05 respectively. The contestant pays \$5 to play the game and is awarded the amount of money indicated by the number where the spinner lands. Is this a fair game?

Solution: By fair, it is meant that the contestant should have an expected return equal to the price she pays to play the game. To answer the question, the expected (average) winnings from playing the game need to be computed. Let the random variable X represent the player's winnings.

$$E[X] = \sum_{x} x \cdot p(x) = (1 \times 0.50) + (5 \times 0.45) + (30 \times 0.05) = 4.25$$

Therefore, this game is not fair, as the house makes an average of 75 cents each time the game is played.

Another interpretation of the expected value of the random variable X is to view it as a weighted mean. Code to compute the expected value using (3.4) and using the function `weighted.mean()` are given next.

```
> x <- c(1,5,30)
> px <- c(0.5,0.45,0.05)
> EX <- sum(x*px)
> WM <- weighted.mean(x,px)
> c(EX,WM)
[1] 4.25 4.25
```

Often, a random variable itself is not of interest but rather some function of it is important, say $g(X)$, of the random variable X. The

expected value of a function $g(X)$ of the random variable X with **pdf** $p(x)$ is

$$E\big[g(X)\big] = \sum_x g(x) \cdot p(x). \tag{3.5}$$

Example 3.19 Consider Example 3.18 for which the random variable Y is defined to be the player's net return. That is $Y = X - 5$ since the player spends \$5 to play the game. What is the expected value of Y?

Solution: The expected value of Y is

$$E[Y] = \sum_x (x-5)\cdot p(x) = (-4\times 0.50)+(0\times 0.45)+(25\times 0.05) = -0.75$$

To compute the answer with S use

```
> x <- c(1,5,30)
> px <- c(0.5,0.45,0.05)
> EgX <- sum((x-5)*px)
> WgM <- weighted.mean((x-5),px)
> c(EgX,WgM)
[1] -0.75 -0.75
```

Rules of Expected Value The function $g(X)$ is often a linear function $a + bX$, where a and b are constants. When this occurs, $E[g(X)]$ is easily computed from $E[X]$. In Example 3.19 a and b were -5 and 1 respectively for the linear function $g(X)$. The following rules for expected value, when working with a random variable X and constants a and b, are true.

1. $E[bX] = bE[X]$

2. $E[a + bX] = a + bE[X]$

Unfortunately, if $g(X)$ is not a linear function of X, such as $g(X) = X^2$, the $E\left[X^2\right] \neq (E[X])^2$. In general, $E\big[g(X)\big] \neq g\big(E[X]\big)$.

3.4.4 Moments

Another way to define the expected value of a random variable is with **moments**. However, knowing the mean (expected value) of a distribution does not tell the whole story. Several distributions may have the same mean. In this case, additional information, such as the spread of the distribution and the symmetry of the distribution, is helpful in distinguishing among various distributions.

The $\boldsymbol{r}^{\text{th}}$ **moment about the origin** of a random variable X, denoted α_r, is defined as $E\left[X^r\right]$. Note that $\alpha_1 = E\left[X^1\right]$ is called the mean of the distribution of X, also denoted μ_X or simply μ. The special moments defined next are important in the field of statistics as they help describe a random variable's distributional shape.

The $\boldsymbol{r}^{\text{th}}$ **moment about the mean** of a random variable X, denoted μ_r, is the expected value of $(X - \mu)^r$. However, all moments do not exist.

For the r^{th} moment about the origin of a discrete random variable to be well defined, $\sum_{i=1}^{\infty} |x_i^r| \mathbb{P}(X = x_i)$ must be less than ∞.

The **r^{th} moment about the origin** of a random variable X, denoted α_r, is defined as $E[X^r]$.

$$\text{Moments about } 0$$
$$E[X^r] = \alpha_r$$

The **r^{th} moment about the mean** of a random variable X, denoted μ_r, is the expected value of $(X-\mu)^r$.

$$\text{Moments about } \mu$$
$$E[(X-\mu)^r] = \mu_r$$

3.3.4.1 Variance The second moment about the mean is called the **variance** of the distribution of X, or simply the variance of X.

$$\text{var}[X] = \sigma_X^2 = E\left[(X-\mu)^2\right] = E\left[X^2\right] - \mu^2 \qquad (3.7)$$

The positive square root of the variance is called the **standard deviation**, and is denoted σ_X. The units of measurement for standard deviation are always the same as the those for the random variable X. One way to avoid this unit dependency is to use the **coefficient of variation**, a unitless measure of variability.

DEFINITION 3.1: **Coefficient of variation** — When $E[X] \neq 0$,

$$CV_X = \frac{\sigma_X}{|E[X]|}, \qquad (3.8)$$

MARQUETTE UNIVERSITY
Be The Difference.
65

3.4.4.2 Rules of Variance If X is a random variable with mean μ and a and b are constants, then

1. $\text{var}[b] = 0$
2. $\text{var}[aX] = a^2\text{var}[X]$
3. $\text{var}[aX + b] = a^2\text{var}[X]$

Note that once $\text{var}[aX + b] = a^2\text{var}[X]$ is proved, $\text{var}[b] = 0$ and $\text{var}[aX] = a^2\text{var}[X]$ have been implicitly shown.

Proof:

$$\begin{aligned}\text{var}[aX + b] &= E[(aX + b - E[aX + b])^2] = E[(aX + b - a\mu - b)^2] \\ &= E[(aX - a\mu)^2] = a^2 E\left[(X - \mu)^2\right] = a^2\text{var}[X].\end{aligned}$$

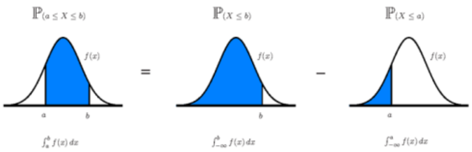
42

MARQUETTE UNIVERSITY
Be The Difference.

3.4.5 Continuous Random Variables Recall that discrete random variables could only assume a countable number of outcomes. When a random variable has a set of possible values that is an entire interval of numbers, X is a **continuous random variable**. For example, if 12 ounce can of beer is randomly selected and its actual fluid contents X is measured, then X is a continuous random variable because any value for X between 0 and the capacity of the beer can is possible.

The function $f(x)$ is a **pdf** for the continuous random variable X, defined over the set of real numbers $\mathbb{R}$ if,

1. $f(x) \geq 0, \ -\infty < x < \infty.$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\, dx = 1.$
3. $\mathbb{P}(a \leq X \leq b) = \displaystyle\int_a^b f(x)\, dx.$



43

DEFINITION 3.2: **Cumulative Distribution Function** — The **cdf**, $F(x)$, of a continuous random variable X with **pdf** $f(x)$ is

$$F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f(t)\,dt, \quad -\infty < x < \infty \qquad (3.10)$$

According to Definition 3.2, the **cdf** is derived from an existing **pdf**. Further, according to the fundamental theorem of calculus, the other direction is also true since $F'(x) = f(x)$ for all values of x for which the derivative $F'(x)$ exists.

Continuous **cdf**s have the following properties:

1. $0 \le F(x) \le 1$.
2. If $a < b$, then $F(a) \le F(b)$ for any real numbers a and b. In other words, $F(x)$ is a nondecreasing function of x.
3. $\lim_{x \to \infty} F(x) = 1$
4. $\lim_{x \to -\infty} F(x) = 0$

Example 3.20 ▷ ***Calculations of* pdf *and* cdf** ◁ Suppose X is a continuous random variable with **pdf** $f(x)$ where

$$f(x) = \begin{cases} k(1-x^2) & \text{if } -1 < x \le 1, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Find the constant k so that $f(x)$ is a **pdf** of the random variable X.

(b) Find the **cdf** for X.

(c) Compute $\mathbb{P}(-0.5 \le X \le 1)$.

(d) Graph the **pdf** and **cdf** of X with S.

Solution: The answers are:

(a) Using property 2 from Box (3.9) for the **pdf** of a continuous random variable write

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-1}^{1} k(1-x^2)\,dx$$
$$= k\left[x - \frac{x^3}{3}\right]\Big|_{-1}^{1} = k\left[\left(1-\frac{1}{3}\right) - \left(-1-\frac{-1}{3}\right)\right]$$
$$= k\left[\frac{2}{3} - \frac{-2}{3}\right] = k\frac{4}{3} \Rightarrow k = \frac{3}{4}.$$

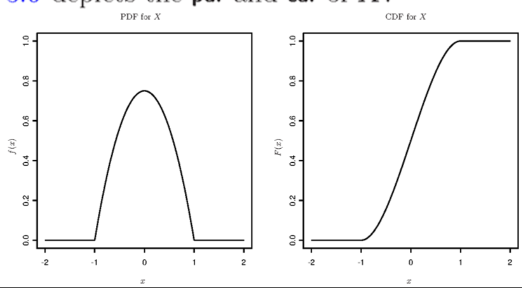
(b)
$$F(x) = \begin{cases} & \text{if } x \le -1 \\ \int_{-1}^{x} \frac{3}{4}(1-t^2)dt = \frac{-x^3}{4} + \frac{3x}{4} + \frac{1}{2} & \text{if } -1 < x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$$

46

(c) Using property 3 from Box (3.9) for the **pdf** of a continuous random variable, write

$$\mathbb{P}(-0.5 \le X \le 1) = F(1) - F(-0.5)$$
$$= \left(\frac{-1^3}{4} + \frac{3\cdot 1}{4} + \frac{1}{2}\right) - \left(-\frac{(\frac{-1}{2})^3}{4} + \frac{3\cdot -\frac{1}{2}}{4} + \frac{1}{2}\right)$$
$$= \left(\frac{-1}{4} + \frac{3}{4} + \frac{1}{2}\right) - \left(\frac{1}{32} + \frac{-3}{8} + \frac{1}{2}\right)$$
$$= 1 - \frac{5}{32} = \frac{27}{32} = 0.84375.$$

(d) Figure 3.6 depicts the **pdf** and **cdf** of X.



47

CODE:

```
➢ par(mfrow=c(1,2), pty="s")
➢ x<-seq(-1,1,0.01)
➢ y<-3/4*(1-x^2)
➢ plot(x, y, xlim=c(-2,2), ylim=c(0,1), type="l",
      xlab="x", ylab="f(x)")
➢ segments(-2,0,-1,0)
➢ segments(1,0,2,0)
➢ title("PDF for X")
➢ y<- -x^3/4 +3*x/4+1/2
➢ plot(x, y, xlim=c(-2,2), ylim=c(0,1), type="l",
      xlab="x", ylab="F(x)")
➢ segments(-2,0,-1,0)
➢ segments(1,1,2,1)
➢ title("CDF for X")
```

48

3.4.5.1 Numerical Integration with S

The S function `integrate()` approximates the integral of functions of one variable over a finite or infinite interval and estimates the absolute error in the approximation.

$$\mathbb{P}(-0.5 \leq X \leq 1) = \int_{-0.5}^{1} \frac{3}{4}\left(1 - x^2\right)\, dx = \frac{3x}{4} - \frac{x^3}{4}\bigg|_{-0.5}^{1} = 0.84375.$$

The following code computes $\mathbb{P}(-0.5 \leq X \leq 1)$ using the function `integrate()` for R and S-PLUS respectively.

```
➢ fx <- function(x){3/4-3/4*x^2}
➢ integrate(fx, lower=-0.5, upper=1)         # R
0.84375 with absolute error < 9.4e-15

➢ fx <- function(x){3/4-3/4*x^2}
➢ integrate(fx, lower=-0.5, upper=1)$integral       # S-PLUS
[1] 0.84375
```

49

MODE - MEDIAN – PERCENTILE

3.4.5.2 Mode, Median, and Percentiles The **mode** of a continuous probability distribution, just like the mode of a discrete probability distribution, is the x value most likely to occur. If more than one such x value exists, the distribution is multimodal.

The **median** of a continuous distribution is the value m such that

$$\int_{-\infty}^{m} f(x)\,dx = \int_{m}^{\infty} f(x)\,dx = \frac{1}{2}.$$

The j^{th} **percentile** of a continuous distribution is the value x_j such that

$$\int_{-\infty}^{x_j} f(x)\,dx = \frac{j}{100}.$$

50

EXAMPLE 3.21

Example 3.21 Given a random variable X with **pdf**

$$f(x) = \begin{cases} 2e^{-2x} & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

(a) Find the median of the distribution.

(b) Find the 25^{th} percentile of the distribution.

Solution: The answers are:

(a) The median is the value m such that $\int_{0}^{m} 2e^{-2x}\,dx = 0.5$ which implies

51

(a) The median is the value m such that $\int_0^m 2e^{-2x}\,dx = 0.5$ which implies

$$\begin{aligned} -e^{-2x}\big|_0^m &= 0.5 \\ -e^{-2m} + 1 &= 0.5 \\ -e^{-2m} &= 0.5 - 1 \\ ln(e^{-2m}) &= ln(0.5) \\ m &= \frac{ln(0.5)}{-2} = 0.3466 \end{aligned}$$

(b) The 25th percentile is the value x_{25} such that $\int_0^{x_{25}} 2e^{-2x}\,dx = 0.25$ which implies

$$\begin{aligned} -e^{-2x}\big|_0^{x_{25}} &= 0.25 \\ -e^{-2x_{25}} + 1 &= 0.25 \\ -e^{-2x_{25}} &= 0.25 - 1 \\ ln(e^{-2x_{25}}) &= ln(0.75) \\ x_{25} &= \frac{ln(0.75)}{-2} = 0.1438 \end{aligned}$$

EXAMPLE 3.22

Example 3.22 Given a random variable X with **pdf**

$$f(x) = \begin{cases} 2\cos(2x) & \text{if } 0 < x < \pi/4 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find the mode of the distribution.

(b) Find the median of the distribution.

(c) Draw the **pdf** and add vertical lines to indicate the values found in part b.

Solution: The answers are:
(a) The function $2\cos 2x$ does not have a maximum in the open interval $(0, \pi/4)$ since the derivative $f'(x) = -4\sin 2x$ does not equal 0 in the open interval $(0, \pi/4)$.

(b) The median is the value m such that

$$\int_0^m 2\cos 2x\,dx = 0.5$$

$$\Downarrow$$

$$\begin{aligned} \sin 2x\big|_0^m = \sin 2m &= 0.5 \\ 2m &= \arcsin(0.5) \\ m &= \frac{\pi}{12} \end{aligned}$$

(c) The R commands used to create Figure 3.7 follow.

```
➢ curve(2*cos(2*x),0,pi/4)
➢ abline(v=pi/12,lty=2,lwd=2)
```

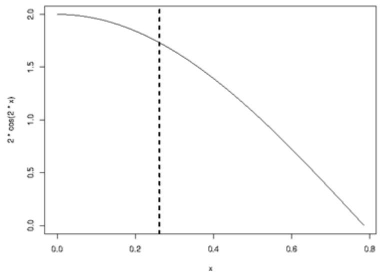


Figure 3.7: Graph of $2\cos(2x)$ from 0 to $\frac{\pi}{4}$ with R

MARQUETTE UNIVERSITY
Be The Difference.

3.4.5.3 Expectation of Continuous Random Variables

For continuous random variables, the definitions associated with the expectation of a random variable X or a function, say $g(X)$, of X are identical to those for discrete random variables except the summations are replaced with integrals and the probability density functions are represented with $f(x)$ instead of $p(x)$. The **expected value** of a continuous random variable X is

$$E[X] = \mu_X = \int_{-\infty}^{\infty} x \cdot f(x)\,dx. \tag{3.12}$$

When the integral in (3.12) does not exist, neither does the expectation of the random variable X. The expected value of a function of X,

MARQUETTE UNIVERSITY
Be The Difference.

89

say $g(X)$, is

$$E\big[g(X)\big] = \int_{-\infty}^{\infty} g(x) \cdot f(x)\, dx. \tag{3.13}$$

Using the definitions for moments about 0 and μ given in (3.6) which relied strictly on expectation in conjunction with (3.13), the **variance** of a continuous random variable X is written as

$$\text{var}[X] = \sigma_X^2 = E\big[(X-\mu)^2\big] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx. \tag{3.14}$$

56

EXAMPLE 3.23

MARQUETTE UNIVERSITY
Be The Difference.

Example 3.23 Given the function

$$f(x) = k, \quad -1 < x < 1$$

of the random variable X

(a) Find the value of k to make $f(x)$ a **pdf**. Use this k for parts (b) through (c).
(b) Find the mean of the distribution using (3.12).
(c) Find the variance of the distribution using (3.14).

Solution: The answers are:

(a) Since $\int_{-\infty}^{\infty} f(x)\, dx$ must equal 1 for $f(x)$ to be a **pdf**, set $\int_{-1}^{1} k\, dx$ equal to one and solve for k.

$$\int_{-1}^{1} k\, dx = 1 \Rightarrow kx\big|_{-1}^{1} = 1$$

$$2k = 1 \Rightarrow k = \frac{1}{2}$$

57

(b) The mean of the distribution using (3.12) is

$$E[X] = \mu_X = \int_{-1}^{1} \frac{1}{2} x \, dx = \left. \frac{x^2}{4} \right|_{-1}^{1} = 0$$

(c) The variance of the distribution using (3.14) is

$$\text{var}[X] = \sigma_X^2 = E\big[(X-\mu)^2\big] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) \, dx = \int_{-1}^{1} (x-0)^2 \frac{1}{2} \, dx = \left. \frac{x^3}{6} \right|_{-1}^{1} = \frac{1}{3}$$

■

3.4.6 Markov's Theorem and Chebyshev's Inequality

Theorem 3.1 Markov's Theorem If X is a random variable and $g(X)$ is a function of X such that $g(X) \geq 0$, then for any positive K

$$\mathbb{P}\big(g(X) \geq K\big) \leq \frac{E[g(X)]}{K}. \tag{3.15}$$

Proof:

Step 1. Let $I\big(g(X)\big)$ be a function such that

$$I\big(g(X)\big) = \begin{cases} 1 & \text{if } g(X) \geq K, \\ 0 & \text{otherwise.} \end{cases}$$

Step 2. Since $g(X) \geq 0$ and $I\big(g(X)\big) \leq 1$, when the first condition of $I\big(g(X)\big)$ is divided by K,

$$I\big(g(X)\big) \leq \frac{g(X)}{K}.$$

MARQUETTE UNIVERSITY Be The Difference.

Step 3. Taking the expected value,

$$E[I(g(X))] \leq \frac{E[g(X)]}{K}.$$

Step 4. Clearly the

$$\begin{aligned} E\left[I\big(g(X)\big)\right] &= \sum_x I\big(g(x)\big) \cdot p(x) \\ &= \left[1 \cdot \mathbb{P}\Big(I\big(g(X)\big) = 1\Big)\right] + \left[0 \cdot \mathbb{P}\Big(I\big(g(X)\big) = 0\Big)\right] \\ &= \left[1 \cdot \mathbb{P}\big(g(X) \geq K\big)\right] + \left[0 \cdot \mathbb{P}\big(g(X) < K\big)\right] \\ &= \mathbb{P}\big(g(X) \geq K\big). \end{aligned}$$

Step 5. Rewriting,

$$\mathbb{P}\big(g(X) \geq K\big) \leq \frac{E[g(X)]}{K}$$

the inequality from (3.15) to be proven.

60

MARQUETTE UNIVERSITY Be The Difference.

If $g(X) = (X-\mu)^2$ and $K = k^2\sigma^2$ in (3.15) it follows that

$$\mathbb{P}\left((X-\mu)^2 \geq k^2\sigma^2\right) \leq \frac{E\left[(X-\mu)^2\right]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}. \qquad (3.16)$$

Working inside the probability on the left side of the inequality in (3.16), note that

$$\begin{aligned} \left((X-\mu)^2 \geq k^2\sigma^2\right) &\Rightarrow \left(X-\mu \geq \sqrt{k^2\sigma^2}\right) \text{ or } \left(X-\mu \leq -\sqrt{k^2\sigma^2}\right) \\ &\Rightarrow \left(|X-\mu| \geq \sqrt{k^2\sigma^2}\right). \end{aligned}$$

Using this, rewrite (3.16) to obtain

$$\mathbb{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}, \qquad (3.17)$$

which is known as **Chebyshev's Inequality**.

DEFINITION 3.3: **Chebyshev's Inequality** — Can be stated as any of

(a) $\mathbb{P}(|X-\mu| \geq k) \leq \frac{\sigma^2}{k^2}$

(b) $\mathbb{P}(|X-\mu| < k) \geq 1 - \frac{\sigma^2}{k^2}$

(c) $\mathbb{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

(d) $\mathbb{P}(|X-\mu| < k\sigma) \geq 1 - \frac{1}{k^2}$

61

3.4.7 Weak Law of Large Numbers

An important application of Chebyshev's Inequality is proving the **Weak Law of Large Numbers**. The Weak Law of Large Numbers provides proof of the notion that if n independent and identically distributed random variables, $X_1, X_2, \ldots, X_n$ from a distribution with finite variance are observed, then the sample mean, $\overline{X}$, should be very close to μ provided n is large. Mathematically, the Weak Law of Large Numbers states that if n independent and identically distributed random variables, $X_1, X_2, \ldots, X_n$ are observed from a distribution with finite variance, then for all $\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) = 0. \tag{3.18}$$

MARQUETTE UNIVERSITY Be The Difference.

Proof: Consider the random variables $X_1, \ldots, X_n$ such that the mean of each one is μ and the variance of each one is σ^2. Since

$$E\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \mu \quad \text{and} \quad \text{var}\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] = \frac{\sigma^2}{n},$$

$$\text{(a) } \mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

use version (a) of Chebyshev's Inequality with $k = \epsilon$ to write

$$\mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2},$$

which proves (3.18) since

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0.$$

3.4.8 Skewness Earlier it was discussed that the second moment about the mean of a random variable X is the same thing as the variance of X. Now, the third moment about the mean of a random variable X is used in the definition of the skewness of X. To facilitate the notation used with skewness, first define a **standardized** random variable X^* to be:

$$X^* = \frac{X - \mu}{\sigma},$$

where μ is the mean of X and σ is the standard deviation of X. Using the standardized form of X, it is easily shown that $E[X^*] = 0$ and $\text{var}[X^*] = 1$. Define the skewness of a random variable X, denoted γ_1, to be the third moment about the origin of X^*.

$$\gamma_1 = E\left[(X^*)^3\right] = \frac{E\left[(X-\mu)^3\right]}{\sigma^3} \tag{3.19}$$

Positive values for γ_1 indicate a distribution that is skewed to the right while negative values for γ_1 indicate a distribution that is skewed to the left. If the distribution of X is symmetric with respect to its mean, then its skewness is zero. That is, $\gamma_1 = 0$ for distributions which are symmetric about their mean. Examples of distributions with various γ_1 coefficients are shown in Figure 3.8.

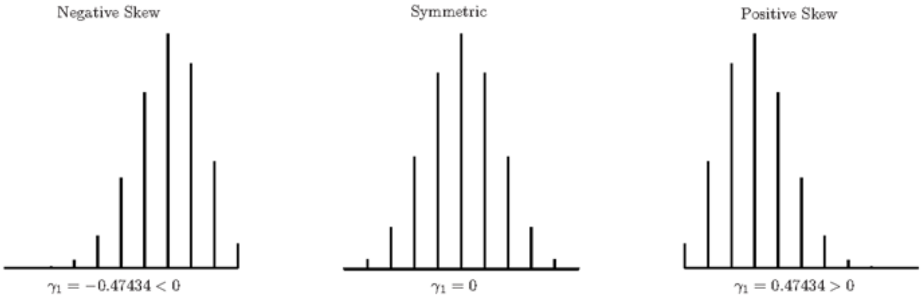


Figure 3.8: Distributions with γ_1 (skewness) coefficients that are negative, zero, and positive respectively.

MARQUETTE UNIVERSITY Be The Difference.

Example 3.25 Let the **pdf** of X be defined by $p(x) = x/15$, $x = 1, 2, 3, 4, 5$. Compute γ_1 for the given distribution.

Solution: The value of γ_1 is computed to be

$$\gamma_1 = E[(X^*)^3] = \frac{E[(X-\mu)^3]}{\sigma^3} = -0.5879747$$

which means the distribution has a negative skew. To compute the answer with S, the following facts are used:

1. $\mu = E[X]$
2. $\sigma = \sqrt{E[X^2] - E[X]^2}$
3. $X^* = \frac{X-\mu}{\sigma}$
4. $\gamma_1 = E\left[(X^*)^3\right]$

```
x <- 1:5
px <- x/15
EX <- sum(x*px)
sigmaX <- sqrt(sum(x^2*px) - EX^2)
X.star <- (x-EX)^3/sigmaX^3
skew <- sum(X.star*px)
skew
[1] -0.5879747
```


MARQUETTE UNIVERSITY Be The Difference.

3.4.9 Moment Generating Functions

Finding the first, second, and higher moments about the origin using the definition $\alpha_r = E[X^r]$ is not always an easy task. However, one may define a function of a real variable t called the moment generating function, **mgf**, that can be used to find moments with relative ease provided the **mgf** exists. Given a random variable X with **pdf** $p(x)$, the **mgf** of X, written $M_X(t)$, is defined as

$$M_X(t) = E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx, \quad -h < t < h. \qquad (3.20)$$

The moment generating function (MGF) of a discrete PMF is

$$M_X(t) = E(e^{tX}) = \sum_x e^{tx} f(x)$$

The MGF does not always exist!

When the MGF exists, it is unique and completely determines the distribution.

MARQUETTE UNIVERSITY
Be The Difference.

Example 3.26 Given the function

$$f(x) = k, \quad -1 < x < 1$$

of the random variable X, find the **mgf** of the distribution using (3.20).

Solution: The reader may verify that a value of $k = \frac{1}{2}$ produces a valid **pdf**. The **mgf** of the distribution will then be

$$\begin{aligned} M_X(t) = E\left[e^{tX}\right] &= \int_{-\infty}^{\infty} e^{tx} f(x)\,dx, \quad -h < t < h \\ &= \int_{-1}^{1} e^{tx} \frac{1}{2}\,dx = \left.\frac{e^{tx}}{2t}\right|_{-1}^{1} \\ &= \frac{e^{t} - e^{-t}}{2t}, \quad t \neq 0. \end{aligned}$$

Note that if $t = 0$, then $M_X(t) = 1$ since $M_X(t) = E\left[e^{tX}\right] = E\left[e^{0}\right] = 1$. Therefore, the **mgf** is written

$$M_X(t) = \begin{cases} \dfrac{e^{t} - e^{-t}}{2t} & \text{if } t \neq 0 \\ 1 & \text{if } t = 0. \end{cases}$$

MARQUETTE UNIVERSITY
Be The Difference.

Theorem 3.2 If X has **mgf** $M_X(t)$, then the derivatives of $M_X(t)$ of all orders exist at $t = 0$, and

$$E[X^n] = \frac{d^n}{dt^n} M_X(t)|_{t=0}.$$

- **Sketch of the Proof (for Discrete R.V.):**

The MGF can be written as

$$\begin{aligned} M(t) &= \sum_{j=1}^{\infty} e^{tx_j} f(x_j) \\ &= \sum_{j=1}^{\infty} f(x_j) \sum_{k=0}^{\infty} \frac{(tx_j)^k}{k!} \\ &= \sum_{k=0}^{\infty} \sum_{j=1}^{\infty} \frac{(tx_j)^k}{k!} f(x_j) \\ &= \sum_{k=0}^{\infty} \frac{(t)^k}{k!} E\left(X^k\right) \end{aligned}$$

Theorem 3.2 If X has mgf $M_X(t)$, then the derivatives of $M_X(t)$ of all orders exist at $t = 0$, and

$$E[X^n] = \frac{d^n}{dt^n} M_X(t)|_{t=0}.$$

- **Sketch of the Proof (for Discrete R.V.):**

If we differentiate the MGF w.r.t. t and evaluate at t=0,

$$\begin{aligned}
\frac{\partial^n}{\partial t^n} M(t) &= \frac{\partial^n}{\partial t^n} \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} \frac{(tx_j)^k}{k!} f(x_j) \\
&= \sum_{j=1}^{\infty} \sum_{k=0}^{\infty} f(x_j) \frac{\partial^n}{\partial t^n} \frac{(tx_j)^k}{k!} \\
&= \sum_{j=1}^{\infty} f(x_j) \frac{\partial^n}{\partial t^n} \left(1 + \frac{(tx_j)}{1!} + \frac{(tx_j)^2}{2!} + \ldots + \frac{(tx_j)^n}{n!} + \ldots \right) \\
&= \sum_{j=1}^{\infty} f(x_j) \left(x_j^n + t x_j^{n+1} \ldots \right) \quad \text{t=0} \longrightarrow E(X^n)
\end{aligned}$$

QUESTIONS?

- **ANY QUESTION?**