Chapter 5

§5.2 Random Variables, Probability Histograms and Probability Distributions

A <u>random variable</u> (r.v.) can be either continuous or discrete. It takes on the possible values of an experiment. It is usually denoted: x when discussing values
X when describing the outcomes

Example: a)   What are the values x can take on for the roll of a single die? Is this a discrete or continuous r.v.?

b)   What are the values x can take on for the altitude of an airplane that takes off from San Francisco airport, assuming that it does not crash? Is this a discrete or continuous r.v.?

c)   Suppose a coin is tossed twice so that the sample space, S, is the following set {HH, HT, TT, TH}. Let X represent the number of heads that come up. What are the possible values of the r.v. x?

A <u>probability distribution</u> is a function that describes the probability associated with each value of a random variable.

Example: Describe the probability distribution associated with the rolls of a single die. In order to find the distribution you must first find the possible values of the r.v. x and then based upon classic probability you will find the probability of obtaining those values.

**Note: There is a functional relationship between the values of a r.v. and the probabilities associated with them.

Example: Describe the probability distribution associated with the number of heads obtained when a coin is tossed twice.

We should now discuss the fact that associated with every probability distribution there are certain rules which must be adhered to. They are as follows:

      1. $\sum P(x) = 1$
      2. $0 \le P(x) \le 1$

Example: Notice in the rolled die experiment where $X = \#$ on the die
      that the $\sum P(x) = 1$ and that $0 \le P(x) \le 1$

Your Turn: Show that the $\sum P(x) = 1$ and that $0 \le P(x) \le 1$
holds true for the coin example

Example: Determine if the following is a probability distribution and indicate which of the rules have been violated if it is not a pdf.

a)

x	1	2	3
f(x)	¼	¼	¾

b)

x	0	1	2
f(x)	- ¼	1	0

A **probability histogram** is a special case of a relative frequency histogram. It shows each probability as a rectangle whose area is equivalent to the probability of the random variable's value. As such the area under the "curve" is always equal to one. The following is a summary of the characteristics of a probability histogram.

1. Every bar is centered over a value of a random variable.
2. Every bar is one unit wide.
3. Every bar touches the one next to it.
4. The height of a bar is equivalent to the probability of the r.v.'s value
5. The area in each bar is equivalent to the probability (2 & 4 multiplied)
6. The sum of the areas under each bar is always 1. Said another way the area under the "curve" is always one.

   Example: Draw a probability histogram for the die rolling example.

   Your Turn: Draw a probability histogram for the coin toss example.

From a probability distribution we can see 3 characteristics of the data:
1) Shape – Via the histogram
2) Mean – By calculation or by the histogram
3) Variance – By calculation or by the histogram

Finding the Mean Using the PDF
$$\mu = \sum x \cdot P(x)$$

Finding the Variance Using the PDF
$$\sigma^2 = [\sum x^2 \cdot P(x)] - \mu^2 \qquad \text{or} \qquad \sigma^2 \ \sum[(x - \mu)^2] \bullet P(x)$$

Note: The first formula, although it looks complicated, is the easier formula.

   Example: Find the shape, mean and standard deviation of the coin flipping example.

Your Turn: Find the shape, mean and standard deviation of the die example.

The last item of discussion is the **expected value**. The **expected value** of a distribution is the same as the mean of the distribution. It can be used to determine net profits and losses, etc and comes from a science called **decision theory**. The expected value has its own notation: **E(x)**

> **Example:** A game is to be played with a single dice that is assumed to be fair. In the game, a player wins $20 if a 2 turns up, $40 if a 4 turns up, loses $30 if a 6 turns up and neither wins nor loses ($0) if any other number turns up. Follow the step by step process of finding the expected profit in playing such a game, by answering each of the following questions.
> a) Give the PDF for rolling a single dice.

> b) Use the PDF in a) to create the PDF for our game.
> Hint: X =Winnings and the probabilities derive from a)

> c) What is the expected profit, in the long run, from playing this game?

> **Your Turn:** In a lottery there are 200 prizes worth $5, 20 prizes worth $25, and 5 prizes worth $100. If there are 10,000 tickets sold, what is the expected winnings for this lottery? BTW, the expected winnings would be considered the expected price to pay for the ticket!

Example: For the following scenario, using a factor tree as a method of finding probabilities, find the expected number of defective radios in a batch.

The probability of a radio produced by the Acme Radio Company being defective is $^1/_2$. Give the PDF for the number of defective radios in a batch of 4.

Note: *We can also use counting techniques to arrive at these probabilities. There number of ways to get a 0 defectives in 4, or 1 defective in 4 or 2 defectives in 4 or 3 defectives in 4 or 4 defectives in 4 can be found using a combination since it is a combination of r items from 4, where order doesn't matter. It doesn't matter for instance if we get one defective in 4 from the 1st, 2nd, 3rd or 4th try, just that we got one defective. Then based upon the multiplication rule; there are 2 possible outcomes from 4 choices, $2 \cdot 2 \cdot 2 \cdot 2 = 16$ possible outcomes. The probability of each outcome therefore is defined by classic probability, P(1) $=_4C_1/16$.*

§5.3 The Binomial Distribution

<div style="border:1px solid black; padding:10px;">

Characteristics of a Binomial Distribution

1) There are a fixed number of trials.

2) Each trial is independent

3) Outcomes can be classified into only 2 categories – success or failure.

4) Probability for success in each trial remains constant.

</div>

The following is the **Binomial Distribution**

$$P(x) = \frac{n!}{x!(n-x)!} \; p^r \cdot q^{n-r}$$

where

$n = $ # of trials
$x = $ # of successes;
$x = 0, 1, 2, 3, \ldots, n$
$p = $ prob. of success
$q = 1 - p$
quotient is $_nC_r$

We have several ways that we can calculate binomial probabilities using our TI-83/84 calculators. Let's take a moment to cover those methods:

1) Using MATH➜PRB menus Enter: "n" MATH➜PRB➜$_nC_r$ ↵ "r" to find the number of combinations, $_nC_r$, and multiply it by the probability of a success and the probability of a failure to their respective exponents.

2) Using DISTR ➜ binompdf(↵ and input the "n", "p", "x" values ENTER This will give all the probability for a particular x.

3) If you want the entire PDF you can eliminate the "x" and the calculator will return an array of all the probabilities for x = 0, 1, 2,…, n. You can then STO➜L$_{1-6}$ to view all probabilities.

Example: Our last example with the defective radios in a batch of 4 was a binomially distributed random variable.
The probability of a radio produced by the Acme Radio Company being defective is $^1/_2$. Give the PDF for the number of defective radios in a batch of 4.

a) Verify that the above 4 conditions for a binomially distributed random variable have been met.

b)   What are n, p and q?

c)   Use your TI-83/84 to verify the pdf that we so painstakingly
        produced using a tree diagram. Store it in a register for easy
        retrieval.

The TI-83/84 will also give us cumulative probabilities. These probabilities represent the
summed probabilities of the random variable given and those below: $P(X \leq x)$ and if we
want the probability of being a random variable or greater: $P(x \geq x)$ we can also find that
by using $1 - P(X \leq x - 1)$. To find a cumulative probability using your TI-83/84, use the
DISTR➔binocdf(↵ enter "n", "p", "x"↵ and the TI will output the probability of being
less than or equal to "x".

d)   Find the probability of getting 2 or fewer defective in a batch using
        the TI-83/84's binocdf function.

e)   Using the binomial pdf stored in your Stat editior, add up the
        $P(0) + P(1) + P(2)$ and verify that it is the same as part d).

f)   Find the probability of getting more than 2 defectives in a batch
        using the TI-83/84's binocdf function.

Note: This is 1 – P(X≤2). You can literally use the 1 – binocdf(n, p, x

e)   Using the binomial pdf stored in your Stat editior, add up the
        $P(3) + P(4)$ and verify that it is the same as part f).

Your Turn: Calculate the probability of getting exactly 2 heads in 9 flips of a fair coin. Answer each of the following questions related to this problem.

a)   Verify that all 4 conditions for a binomial probability distribution are met.

b)   List n, p, q and using P(x) notation, the probability that you wish to find.

c)   Using the defining PDF for the binomial, write the P(x). I don't expect you to find it, just to be able to write it into the defining formula.

d)   Calculate P(x) using your TI-83/84.

e)   Using your TI-83/84 calculate the probability of getting at most 3 heads in 9 flips of a fair coin. Use P(x) notation to denote the probability.

f)   Using your TI-83/84 calculate the probability of getting fewer than 4 heads in 9 flips of a fair coin. Use P(x) notation to denote the probability.

g)   Using your TI-83/84 calculate the probability of getting at least 4 heads in 9 flips of a fair coin. Use P(x) notation to denote the probability.

Note: *At least 4 is 4 or more, so we have to do 1 – probability of being 3 or fewer! ;-)*

        h)   Using your TI-83/84 calculate the probability of getting more than 5 heads in 9 flips of a fair coin. Use P(x) notation to denote the probability.

Note: *This is the complement of being at most 5!*

Another method of finding binomial probabilities is using a Binomial Probability Table. Our table is in Appendix II. You will find n's is the left column and all values of X in the column next to the n. Along the top of the table you will find a range of values for p. Under each value of p, you will find the probability of getting the r.v. in the x column. This table is limited because every value of p is not listed and most tables in texts only go to n = 15.

       Example: Use the table to answer the question following this example.
                    A social scientist claims that 60% of all high school seniors capable of doing college work actually attend college. Assuming that his claim is true, find the following:

        a)   The probability that exactly 10 go to college. Use P(x) notation to denote the probability.

        b)   The probability that at least 10 will go to college. Use P(x) notation to denote the probability.

        c)   The probability that at most 8 will go to college. Use P(x) notation to denote the probability.

       Example: A student comes unprepared to an exam, so they will be guessing on each and every question on the multiple choice exam. Each of the 10 problems on the exam has answers choices, a) through e). Answer the following questions based on this scenario.
        a)   What is the probability of the student getting a C or better? Use P(x) notation to denote the probability.

b)   What is the probability that the student will get an A? Use P(x) notation to denote the probability.

A question that might arise based on this scenario, might be:
What would you expect the student to get on the exam?

In order to answer this question, we need to know how to find the **mean** of the Binomial Distribution.

Further questions could also arise, such as:
Between what scores would you consistently expect such a student to score?

In order to answer this question we would need the **standard deviation**. Not to mention that we would have to "define" what we mean by "consistently". One way of defining "consistently", may be within 1 standard deviation of the mean, just as a way of defining "usually", may be within 2 standard deviations and thus "unusual" may be beyond 2 standard deviations. Your book considers an "potential outlier" to be beyond 2.5 standard deviations.

Mean & Standard Deviation of Binomial Distribution

$$\mu = n \cdot p$$

$$\sigma^2 = n \cdot p \cdot q \ \text{ and } \therefore \quad \sigma = \sqrt{n \cdot p \cdot q}$$

Example: So, back to our unprepared student in the last example.
a)   What would you expect the student to get on the test?

b)   In what range would you expect such a student to "consistently" score?

c)   Above what score would you consider to be unusually high for such a student?

One last example, which comes back to our expected value, involves expected cost. Finding expected cost is a means by which a company may decide upon the warranty cost for an item. Expected cost is found by multiplying the expected value by the cost to replace/repair a single item.

> **Example:** A company produces widgets in very large lots, with a defective rate of 10% per lot. For a particular sale, of a batch 4 items, find the expected repair cost if the repair of a single widgit would cost $15.