Math107: An Introduction to Random Variables

Dr. Richard Mikula

Fall 2009

At this point we revisit the realm of probability. To start, it might be helpful to recall some terminology that we discussed earlier.

Given an experiment with outcomes listed in a set S, our sample space, the probability of an event E, where E is a collection of some outcomes in S, is denoted by

$$P(E),$$

and is a number between 0 and 1. The case where

$$P(E) = 0$$

corresponds with the situation where E will most certainly not occur. The case where

$$P(E) = 1$$

corresponds with the situation where E will most certainly occur.

How close to $P(E)$ is to zero or to one is a gauge of how likely the event E will occur.

A **random variable**

$$X$$

is a numerical quantity that is associated with each outcome of an experiment. That is for each outcome O of an experiment, we associate with this outcome a number, which we call X.*

*For those of you that may have heard the term **function** before, a random variable is a numerical valued function, defined on the sample space S of an experiment.

Let us consider several examples:

- Let our experiment consist of rolling a fair six sided die and noting the outcome. Then

$$S = \{1, 2, 3, 4, 5, 6\}.$$

  We may define a random variable X to be the value of the face of the die.

- Let our experiment consist of flipping a fair coin three consecutive times and recording the sequence of $H's$ and $T's$. Thus

$$S = \{HHH, HHT, HTH, THH$$

$$HTT, THT, TTH, TTT\}.$$

  Let us define a random variable X to be the number of $H's$.

- Let an experiment consist of drawing a card from a standard deck of 52 cards, and noting which card was drawn. Thus S is the possible cards in a standard deck of cards. Define a random variable X for this experiment to be the value of the card if it is a number card. Let the random variable be 11 for a jack, 12 for a queen, 13 for a king and 1 for an ace.

If our sample space S consists of outcomes which we generally denote by O. The notation

$$X(O)$$

is just the value for the random variable for the particular outcome O. Considering our third example above

$$X(Ace) = 1.$$

For our second example above

$$X(HTH) = 2.$$

We let x denote a possible value for a given random variable X. We call a random variable **discrete** if its possible values x lie in a discrete set. That is they lie in a set that can be listed explicitly*, whether it be a finite or infinite list.

*that is lie in a finite set, or can be put in a one to one correspondence with the counting numbers $\{1, 2, 3, 4, \cdots\}$

We call a random variable **continuous** if its possible values x lie in an interval or collection of intervals*.

*and can be any value in this interval or collection of intervals

For our three examples above, our random variables are all discrete. To see this note that in the first example, x can be any number in the set

$$\{1, 2, 3, 4, 5, 6\}.$$

In the second example, x can any number in the set

$$\{0, 1, 2, 3\}.$$

In the third example, x can be any number in

$$\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}.$$

To come up with some examples of continuous random variables, we need to be a bit more creative. For the following list of experiments, I will give some examples of continuous random variables:

- Let an experiment consist of going to an elevator, and timing how long it takes for the elevator to arrive. We define X to be the time it takes to arrive.

- Let an experiment be randomly selecting a student from Lock Haven University. Thus S is the collection of students. We define X to be the students height.

Thus, examples of random variables that are times, lengths or temperatures are examples of continuous random variables.

From the probabilities we have associated with the outcomes of an experiment we can calculate the probability of the quantity

$$P(X = x)$$

where X is a random variable defined on S and x is a possible value of X. It is common to write $f(x)$ in place of $P(X = x)$. We commonly call the collection of quantities $f(x)$, when viewed over all the possible values of x, a **probability distribution of** X. If the random variable is discrete, we may commonly list this in a table. For instance, recall our example of flipping a fair coin three times and recording the outcomes, we have

x	$f(x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

Let us consider another example. Recall the experiment of rolling 2 fair six sided dice, and noting the values on each face. The sample space S is the collection of outcomes:

$$(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)$$

$$(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)$$

$$(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)$$

$$(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)$$

$$(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)$$

$$(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)$$

where (a,b) represents the outcome where the value of the first die is a and the value of the second die is b.

Let us define a random variable X for this experiment to be the sum of the two faces values, that is

$$X((a, b)) = a + b.$$

For example,

$$X((1, 1)) = 2, \quad X((4, 3)) = 7.$$

Thus we see the possible values for x are

$$\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

The probability distribution for X is given by the table:

x	$f(x)$
2	$\frac{1}{36}$
3	$\frac{2}{36}$
4	$\frac{3}{36}$
5	$\frac{4}{36}$
6	$\frac{5}{36}$
7	$\frac{6}{36}$
8	$\frac{5}{36}$
9	$\frac{4}{36}$
10	$\frac{3}{36}$
11	$\frac{2}{36}$
12	$\frac{1}{36}$

The quantity $f(x)$ is usually referred to as the **probability density function** for the random variable X.

Some general properties of probability density function f are:

- $$P(X = x) = f(x)$$

- $$\sum_x f(x) = 1$$

  where here the sum is taken over all possible values of x.

- $$0 \leq f(x) \leq 1$$

  for any value x

When studying an experiment with a discrete random variable X, it is sometimes useful to look at what is called the **probability histogram** or the **density histogram**. This is a histogram that on the horizontal axis we list the values of x, and on the vertical axis we list the corresponding possible values for $f(x)$. This we will discuss further in the lecture.

Mathematical Expectation:

Let S be the sample space for an experiment, and let X be a discrete random variable associated to this experiment.

At this point we define the **expected value** of X to be

$$E(X) = \sum_x x f(x)$$

where the sum is taken over all possible values x that X can take on.

The **expected value** value of a random variable X is a number that is a measure of what you should expect to see on average (for the values of X), if you were to perform the experiment a large number of times. That is, suppose you perform an experiment N times, and you observe the results of the experiment for those N trials to be

$$O_1, O_2, O_3, \cdots, O_N,$$

then for N sufficiently large, we should observe

$$\frac{1}{N} \sum_{k=1}^{N} X(O_k) \approx E(X).$$

Let us return to our first three examples considered, namely:

- Let our experiment consist of rolling a fair six sided die and noting the outcome. Then

$$S = \{1, 2, 3, 4, 5, 6\}.$$

  We may define a random variable X to be the value of the face of the die.

- Let our experiment consist of flipping a fair coin three consecutive times and recording the sequence of $H's$ and $T's$. Thus

$$S = \{HHH, HHT, HTH, THH$$

$$HTT, THT, TTH, TTT\}.$$

  Let us define a random variable X to be the number of $H's$.

- Let an experiment consist of drawing a card from a standard deck of 52 cards, and noting which card was drawn. Thus S is the possible cards in a standard deck of cards. Define a random variable X for this experiment to be the value of the card if it is a number card. Let the random variable be 11 for a jack, 12 for a queen, 13 for a king and 1 for an ace.

For the first example we have

x	$f(x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

Thus

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6}$$

$$+4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6}$$

$$= \frac{21}{6} = 3.5.$$

For our second example we have

x	$f(x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

Thus the expected value of X is

$$E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8}$$

$$= \frac{12}{8} = 1.5$$

For our third example, the possible values for X are

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13$$

For each possible value x, we have $f(x) = \frac{4}{52}$ (Why?)

Thus we see that

$$E(X) = 1 \cdot \frac{4}{52} + 2 \cdot \frac{4}{52} + 3 \cdot \frac{4}{52} + 5 \cdot \frac{4}{52} + 6 \cdot \frac{4}{52}$$

$$+7 \cdot \frac{4}{52} + 8 \cdot \frac{4}{52} + 9 \cdot \frac{4}{52} + 10 \cdot \frac{4}{52} + 11 \cdot \frac{4}{52} + 12 \cdot \frac{4}{52} + 13 \cdot \frac{4}{52}$$

$$= 91 \cdot \frac{4}{52} = 7.$$

A Lottery Game: Let us consider the experiment of playing a lottery game that consists of selecting 6 different numbers from the set

$$\{1, 2, 3, 4, \cdots 48, 49\}.$$

The numbers are then drawn every evening by selecting six balls (without replacement) from a container with 49 balls labeled with the numbers 1 through 49.

Suppose you must pay one dollar to play this game. You win 10,001 dollars if you have exactly 5 of the numbers that appeared in the drawing and one incorrect number on your ticket. Moreover, you win 1,000,001 dollars if you have all six numbers that were drawn on your ticket. Otherwise, you loose.

We may define the random variable X to be the amount of money won when you play this lottery game. If you loose $X = -1$, if you have exactly 5 of the correct numbers $X = 10,000$ and if you have all the correct numbers $X = 1,000,000$. Thus we have

x	$f(x)$
-1	$\frac{1997651}{1997688}$
$10,000$	$\frac{43}{2330636}$
$1,000,000$	$\frac{1}{13983816}$

Hence,

$$E(X) = -1 \cdot \frac{1997651}{1997688} + 10,000 \cdot \frac{43}{2330636}$$
$$+1,000,000 \frac{1}{13983816}$$
$$= -\frac{10403557}{13983816} \approx -0.744$$

Question: Does this game seem fair? That is, would you come out ahead in the long run, if you play this game a large number of times?

HW Problem: A term life insurance policy will pay a beneficiary a certain sum of money upon the death of the policy holder. These policies have premiums that must be paid annually. Suppose a life insurance company sells a 100,000 one-year term policy to an 18-year-old male for A dollars. According to *the National Vital Statistics Report, Vol. 47, No. 28,* the probability the male will survive the year is 0.998789. Compute the expected value of this policy to the insurance company. How big should A be in order for the insurance company to make money?

Answer: $A \geq 121.1$ is necessary.

Measures of Central Tendency for a Random Variable X:

So far we way discussed

$$E(X) = \sum_x x f(x)$$

the so-called mathematical expectation of X or the expected value of X. This is a measure of central tendency of X. There are other measures of central tendency, however we shall not consider any others at this time.

Measures of Dispersion of a Random Variable X:

Suppose that you were to perform N trials of some experiment, and X is a random variable defined for this experiment which has a probability density function $f(x)$. Suppose the observed outcomes on these N trials are the outcomes

$$O_1, O_2, O_3, O_4, \cdots, O_{N-1}, O_N$$

and their corresponding random variable values are

$$X(O_1), X(O_2), \cdots, X(O_N).$$

We observed earlier that the the average value of the collection of numbers

$$X(O_1), X(O_2), \cdots, X(O_N),$$

namely

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} X(O_i),$$

is close to the value $E(X)$ provided N is very large. *

*This is a consequence of the law of large numbers, since this computed average $\bar{x}$ is simply the observed values of $X(O)$ times their relative frequencies for the N trials. Moreover, when N is very large, these relative frequencies are close to the values $f(X(O))$.

For this list of observed values of X

$$X(O_1), X(O_2), X(O_3), \cdots, X(O_{N-1}), X(O_N)$$

we can compute the standard deviation, which is simply

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (X(O_i) - \bar{x})^2}.$$

As N gets very large, this will get close to the following quantity

$$sd(X) = \sqrt{\sum_x (x - E(X))^2 f(x)},$$

with the sum taken over all possible values x of X. We call the quantity $sd(X)$ the **standard deviation of the random variable** X.

Some Examples:

- Let our experiment consist of rolling a fair six sided die and noting the outcome. Then

  $$S = \{1, 2, 3, 4, 5, 6\}.$$

  We may define a random variable X to be the value of the face of the die.

- Let our experiment consist of flipping a fair coin three consecutive times and recording the sequence of $H's$ and $T's$. Thus

  $$S = \{HHH, HHT, HTH, THH$$

  $$HTT, THT, TTH, TTT\}.$$

  Let us define a random variable X to be the number of $H's$.

For these examples, we showed that $E(X)$ equals 3.5 and 1.5 respectively. We now will compute the standard deviations for these experiments' random variables.

For the die example:

x	$f(x)$	$x - E(X)$	$(x - E(X))^2$	$(x - E(X))^2 f(x)$
1	$\frac{1}{6}$	-2.5	6.25	1.042
2	$\frac{1}{6}$	-1.5	2.25	0.375
3	$\frac{1}{6}$	-0.5	0.25	0.042
4	$\frac{1}{6}$	0.5	0.25	0.042
5	$\frac{1}{6}$	1.5	2.25	0.375
6	$\frac{1}{6}$	2.5	6.25	1.042

We sum this right most column to get 2.918. Thus

$$sd(X) \approx \sqrt{2.918} \approx 1.708.^*$$

*Here I use $\approx$ instead of $=$ because I rounded at various stanges of my calculation

For the coin example:

x	$f(x)$	$x - E(X)$	$(x - E(X))^2$	$(x - E(X))^2 f(x)$
0	$\frac{1}{8}$	-1.5	2.25	0.282
1	$\frac{3}{8}$	-0.5	0.25	0.094
2	$\frac{3}{8}$	0.5	0.25	0.094
3	$\frac{1}{8}$	1.5	2.25	0.282

Summing the numbers in the right most column gives 0.752. The standard deviation is

$$sd(X) \approx \sqrt{0.752} \approx 0.867.$$

Combinations:

A **combination** is a distinct collection of objects chosen from a larger collection of objects.

The number of combinations of r objects selected from a set of n objects is denoted by

$$\binom{n}{r}$$

and is simply the number of subsets of cardinality r, of a set with cardinality n.

Let us define the **factorial** for a non-negative integer k, denoted by $k!$, to be 1 if $k = 0$ and otherwise given by

$$k! = k \cdot (k-1) \cdots 3 \cdot 2 \cdot 1.$$

Thus $1! = 1$, $2! = 2$, $3! = 6$, $4! = 24$, etc.

It turns out that

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

The following triangle is referred to as **Pascal's triangle**

$$
\begin{array}{c}
1 \\
1 \quad 1 \\
1 \quad 2 \quad 1 \\
1 \quad 3 \quad 3 \quad 1 \\
1 \quad 4 \quad 6 \quad 4 \quad 1 \\
1 \quad 5 \quad 10 \quad 10 \quad 5 \quad 1 \\
1 \quad 6 \quad 15 \quad 20 \quad 15 \quad 6 \quad 1 \\
\cdots \\
\cdots \\
\binom{n}{0} \quad \binom{n}{1} \quad \binom{n}{2} \quad \cdots \quad \binom{n}{r} \quad \cdots \quad \binom{n}{n-1} \quad \binom{n}{n}
\end{array}
$$

Bernoulli Trials and the Binomial Distribution:

Suppose you have an unfair coin, which has the probability of getting a head H on any one flip being $p = 0.49$.* Suppose you flip the coin 5 times. Then each flip is an independent trial, since the outcome of one flip is in no way related to the outcome of another flip.

What is the probability of observing

$$TTHTH$$

as an outcome?

*and thus the probability of getting a tail T on one flip is $q = 1 - p = 0.51$

The probability of getting

$$TTHTH$$

is given by

$$P(TTHTH) = P(T) \cdot P(T) \cdot P(H) \cdot P(T) \cdot P(H)^*$$

$$= q \cdot q \cdot p \cdot q \cdot p$$

$$= q^3 p^2$$

$$= 0.51^3 \cdot 0.49^2 \approx 0.03185.$$

*by independence

Suppose we wished to know the probability of getting two heads on five flips of this unfair coin. Clearly this must be larger than the probability we just computed (why?).

Let us first list all the outcomes which consist of two $H's$ and three $T's$

$$HHTTT, HTHTT, HTTHT, HTTTH,$$

$$THHTT, THTHT, THTTH,$$

$$TTHHT, TTHTH, TTTHH.$$

The probability of each of these outcomes in this event is the same, and is

$$q^3 p^2 \approx 0.03185.$$

Moreover, there are ten outcomes in this event $\{$ getting two $H's$ $\}$. Thus

$$P(2H's) = 10q^3p^2 \approx 0.3185.$$

Can we compute this probability without listing all the outcomes which have two $H's$ and three $T's$?

Actually, yes. If you look at the event

$$\{2H's \quad in \quad 5 \quad flips\}$$

$$= \{HHTTT, HTHTT, HTTHT, HTTTH,$$

$$THHTT, THTHT, THTTH,$$

$$TTHHT, TTHTH, TTTHH\},$$

you see that I need to pick the two locations for the $H's$, out of the five possible locations; I can do this in

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$$

possible ways. The probability of the event $\{2H's\}$ is then

$$\binom{5}{2}p^2q^3.$$

A **Bernoulli trial** is an experiment which has two possible outcomes. We call one of the possible outcomes a *success* and the other a *failure*. Suppose that the probability of a success occurring is p and the probability of a failure occurring is $q = 1 - p$.

We define a second experiment by: Perform a sequence of n Bernoulli trials. We note here that any two trials in this sequence of trials are independent of one another.

Let us define the random variable X for the second experiment to be the number of successes in the n Bernoulli trials. Then the probability of observing x successes in the n trials is given by

$$P(X = x) = f(x) = \binom{n}{x} p^x q^{n-x}.$$

An experiment (like the one we just defined) which has the probability density function

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

for a random variable X defined for that experiment, is said to have a **Binomial Probability Distribution.**

Example: A multiple choice test consists of 6 questions. Suppose each question has four choices for the answer, namely a,b,c and d. Suppose that a student passes the test, if he or she gets at least 4 correct answers out of 6. Suppose a student who did not go to class all semester shows up to take this test. The student does not know any of the material, and randomly guesses on each question. What is the probability that the student passes this test?

Solution: The student passes if he or she gets either 4, 5 or 6 correct answers. Let X be the number of correct answers out of 6. We know that $p = \frac{1}{4}$ and $q = \frac{3}{4}$ are the probabilities of guessing correctly and incorrectly respectively on any one test question. The the probability of passing is

$$P(X = 4) + P(X = 5) + P(X = 6)$$

$$= \binom{6}{4}(\frac{1}{4})^4(\frac{3}{4})^2 + \binom{6}{5}(\frac{1}{4})^5(\frac{3}{4})^1 + \binom{6}{6}(\frac{1}{4})^6(\frac{3}{4})^0$$

$$= 15(\frac{1}{4})^4(\frac{3}{4})^2 + 6(\frac{1}{4})^5(\frac{3}{4})^1 + (\frac{1}{4})^6$$

$$\approx 0.0376.$$

The **expected value** for the number of success for the **binomial distribution** is given by

$$E(X) = np.$$

Also, the standard deviation is given by

$$sd(X) = \sqrt{npq}.$$

In the first example, where we flip the coin five times, the expected value is

$$E(X) = 5 \cdot 0.49 = 2.45$$

and the standard deviation is

$$sd(X) = \sqrt{5 \cdot 0.49 \cdot 0.51} \approx 1.118.$$

Thus, if we flipped the unfair coin 5 times, and recorded the number of heads, and we did this experiment many times, and computed the average of the number of heads we observed on application of the experiment, we would find out that the average is roughly 2.45.

In the test example, the expectation is

$$E(X) = 6 \cdot 0.25 = 1.5$$

and

$$sd(X) = \sqrt{6 \cdot 0.25 \cdot 0.75} \approx 1.061.$$

Thus, if we had a very large classroom, full of students who all randomly guessed on every question, we recorded the number of correct answers each student got, and took an average, we would get a number which is roughly 1.5.

An instant lottery ticket has a 20 percent chance of being a winning ticket. Suppose a person purchases 7 tickets. What is the probability of having 2 winning tickets?

Here the probability of having a winning ticket is $p = 0.2$. Thus, the probability of having a loosing ticket is $q = 1 - p = 0.8$. If X is the number of winning tickets in our collection of 7 tickets, then

$$P(X = 2) = \binom{7}{2} p^2 q^5$$

$$= \frac{7!}{2!5!}(0.2)^2(0.8)^5$$

$$\approx 0.275.$$

Here $E(X)$ and $sd(X)$ are given by

$$E(X) = 7 \cdot 0.2 = 1.4$$

and

$$sd(X) = \sqrt{7 \cdot 0.2 \cdot 0.8} = \sqrt{1.12} \approx 1.058.$$

Home Work Exercise: Suppose that the probability that a missile hits its target is $p = 0.6$. If 7 missiles are launched at 7 different targets, and the random variable X represents the number of successful hits, find:

1. $P(X = 0)$

2. $P(X = 1, 2, or 3)$

3. $P(X > 3)$

4. $P(X = 7)$

5. $P(X < 7)$

6. The probability that there is at least one hit.

7. $E(X)$

8. $sd(X)$.

Answers: 0.0016384, 0.2881536, 0.710208, 0.0279936, 0.9720064, 0.9983616, 4.2, 1.296148

Chebyshev's Theorem:

The following theorem gives us an upper bound on a probability of a random variable X. If $E(X)$ is the expected value of the random variable, and $sd(X)$ is the standard deviation for the random variable, then for any number k we know:

$$P(E(X) - k \cdot sd(X) < X < E(X) + k \cdot sd(X)) \geq 1 - \frac{1}{k^2}.$$

Example: Suppose a random variable has an expected value of 8 and a standard deviation of 3. Find a lower bound on the probability that X satisfies

$$-4 < X < 20.$$

Solution:

$$P(-4 < X < 20)$$

$$= P(8 - 4 \cdot 3 < X < 8 + 4 \cdot 3) \geq 1 - \frac{1}{4^2} = \frac{15}{16}.$$

Suppose that on a statistics exam the class average is 73 points (out of 100 possible points) and the standard deviation is 10 points. Find a lower bound on the portion of students who have exam scores strictly between 48 and 98.

Solution: Let X be the students exam score, for the experiment that consists of selecting a student from the stat class. Then $E(X) = 73$ and $sd(X) = 10$. We wish to find a lower bound for

$$P(48 < X < 98).$$

To do this we use Chebyshev's theorem, since

$$P(48 < X < 98) = P(73 - 25 < X < 73 + 25)$$

$$= P(73 - 2.5 \cdot 10 < X < 73 + 2.5 \cdot 10) \geq 1 - \frac{1}{2.5^2} = 0.84.$$

Thus at least 84 percent of the students have exam scores strictly between 48 and 98 points.

For the previous example, if say we knew that 48 was the cut-off for a failure, we could definitely conclude that no more than 16 percent of the students failed the exam.

Some Home Work Exercises:

1. A random variable X has $E(X) = 10$ and $sd(X) = 2$. Find lower bounds for:

   (a) $P(7 < X < 13)$

   (b) $P(5 < X < 15)$

   (c) Find an upper bound for $P(|X - 10| \geq 3)$.

2. An electrical firm manufactures a 100-watt light bulb, which, according to specifications written on the package, has a mean life of 900 hours with a standard deviation of 50 hours. You purchase such a light bulb. Find a lower bound on the probability that the light bulb lasts more than 775 hours but less than 1025 hours.

Answers: 1: 0.555555, 0.84, 0.444444; 2: 0.84