5:1a Random Variables

- When a variable can take on any one of several values, which are unpredictable in the short term, the variable is called a random variable.
- Random variables are not haphazard.
- They frequently display an order when large numbers of events are examined
- These numerical measures can be either discrete or continuous.

When a die is rolled there are only a countable number of possible outcomes in the sample space. These are 1, 2, 3, 4, 5, and 6 for a total of 6 outcomes. There are gaps between each possible outcome. We could not get a 6.5, or even a 0 as an outcome. This is the defining characteristic of discrete variables. They are isolated. When we address the issue of probability of these outcomes we will assign relative frequencies to the individual outcomes.

5:1b Random Variables

On the other hand if we are measuring the heights of individuals it is possible to get a height of 70 inches or 70.2 inches.
The smallest increment that we can measure is limited only by the instruments we use to perform the measurement. Theoretically it is possible to find three people whose different heights are in order a, b, c no matter how close a and c are to each other. Variables, which take on any value along a number line, are defined as continuous. When we discuss the probability of these variables we will utilize a probability density function.

Mean of a Discrete Random Variable

Remember that a single die has 6 equally likely outcomes {1,2,3,4,5,6}, if a single die were to be rolled a large number of times (say, for example, 6000) what would the mean of the outcomes.
Since we know that the probability of each outcome is one-sixth and because of the Law of Large Numbers we could write the results as:

5:1c Random Variables

- We would expect that the die would land with a "1" face up 1000 times.

- We would expect that the die would land with a "2" face up 1000 times.

- We would expect that the die would land with a "3" face up 1000 times.

- We would expect that the die would land with a "4" face up 1000 times.

- We would expect that the die would land with a "5" face up 1000 times.

- We would expect that the die would land with a "6" face up 1000 times.

5:1d Random Variables

Then by multiplying the frequency of each outcome by the value of that outcome and summing all of the answers we get;

$1000+2000+3000+4000+5000+6000 = 21000$

Now to find the mean just divide the sum by the number of events and we get $21000/6000 = 3.500$.

5:1e Random Variables

The terminology is that the expected value is 3.5. In general for a discrete random variable the expected value is:

$$\boxed{E(x) = \sum_{i=1}^{n} p(x_i)x_i \ = \mu_x}$$

If the probabilities of the outcomes are not constant but instead are given either by a graph or a table the process is the same.

5:1f Random Variables

In order for a probability distribution to be legitimate or valid the following conditions must be satisfied.

1. $p(x) \geq 0 \textit{ for all } x$

2. $\sum_{i=1}^{n} p(x_i) = 1$

Consider the following situation;

A merchant wants to increase traffic at his grocery store. In order to do this he runs a promotional coupon program. He will issue 4 denominations of coupons, \$1, \$2, \$5, \$10. He prints the tickets with the frequencies given in the table below.

When a customer checks out at the counter he picks a random ticket and the customer gets a discount equal to the amount printed on the ticket. What is the mean discount given?

5:1g Random Variables

Ticket Value	\$1.00	\$2.00	\$5.00	\$10.00
Frequency	.500	.300	.150	.050

After satisfying ourselves that both conditions for a probability distribution are met we can proceed. Are they?

Take a moment and verify!

Applying the formula for the expected value of discrete random variables we get;

$$\mu_x = E(x) = \sum_{i=1}^{4} p(x_i)x_i$$

$$\mu_x = E(x) = (.5)*1 + (.3)*2 + (.15)*5 + (.05)*10$$

$$\mu_x = E(x) = .5 + .6 + .75 + .5 = 2.35$$

In this case the merchant will average a \$2.35 discount for each customer even though no customer will actually get a discount of \$2.35.

5:1h Random Variables

Standard Deviation of a Discrete Random Variable

Recall that the formula for the standard deviation is

$$\sigma = \sqrt{\frac{\sum_{\text{for all i}} (x_i - \mu)^2}{N}}$$

we are going to modify this formula only slightly to include a factor for the probability of each x occurring.

The new formula is

$$\sigma = \sqrt{\sum_{\text{for all i}} (x_i - \mu)^2 \, p(x_i)}$$

note that the

$$(x - \mu)$$

term is just the deviation from the mean as it was in the original formula and the $p(x_i)$ is simply the frequency of this term appearing divided by N.

5:1i Random Variables

Keeping these two pieces of information in mind it is a simple algebra manipulation to convert the original formula in to the new formula.

The variance of the discrete random variable would be:

$$\mathrm{var}\,iance = \sum_{\text{for all i}} \left(x_i - \mu\right)^2 p(x_i) \quad .$$

5:1j Random Variables

Mean and Standard Deviation of a Continuous Random Variable

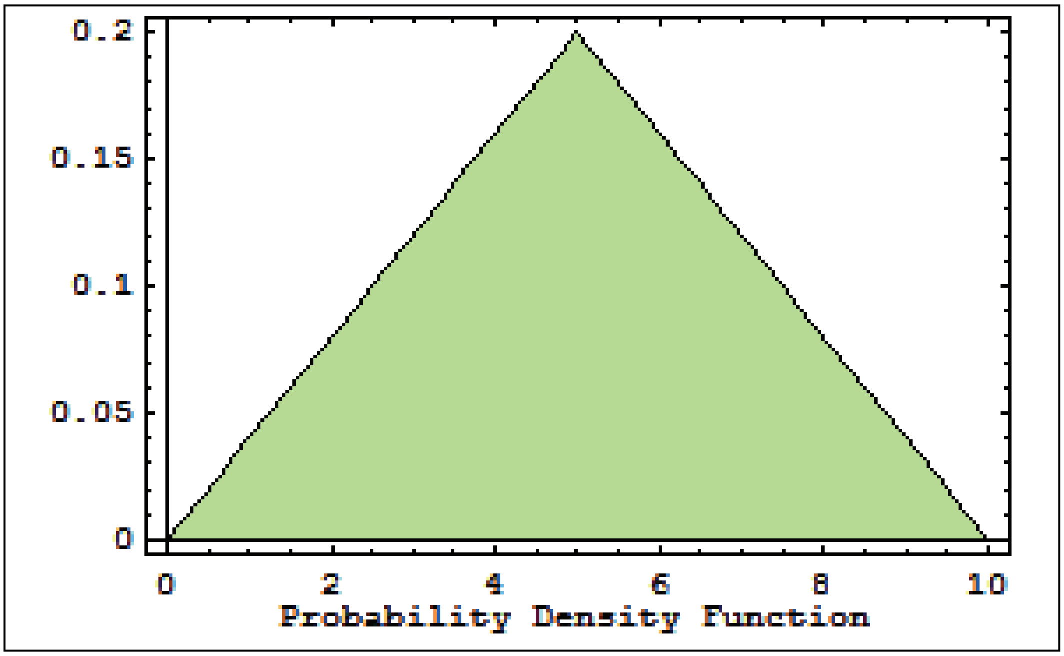
Continuous variables are a little different than discrete variables. The basic idea is the same, namely, assign each possible value a probability, and then find the sum. Unlike discrete variables probabilities can not be assigned to individual values because there are an infinite number of them. We address that problem by shifting from an algebraic analysis to one of a geometric nature.

The probabilities will be considered to be the area under the curve. In order to be a true probability density curve some conditions must be satisfied;

f(x) is a probability density curve if and only if: 1) $f(x) \geq 0 \; for \; all \; x, a \leq x \leq b$ 2) $\text{the area bounded by the curve must} = 1$

5:1k Random Variables

The diagram below is a straightforward graph. Can it be a probability density function?



For each value of x is f(x) >= 0?

Is the area under the curve equal to 1?

Since the answer to both questions is yes this is a valid probability density function.

5:1l Random Variables

What is the probability of the outcome value (x=3)? That's a little difficult to answer, since we spoke of the probability of a continuous random variable as being the area under the curve at that point. Area requires that we multiply a width by a length. In this case there is no width! Calculus gives us tools that enable us to calculate the area under a curve or for regions within that area. We will not discuss those tools here but instead rely on either technology or tables to provide us with the answer.

5:1m Random Variables

The important consideration is that the mean and standard deviation have the same meaning and role as they do for discrete variables. The mean will be the point at which sum of the “moment of torque” for each data point will equal zero. In other words the graph of the distribution would balance at that point. The standard deviation will still be a measure of variability or spread. The greater the standard deviation the more spread out from the mean the data points will be. A small standard deviation will indicate that the data points are clustered relatively close to the mean.

Question: Given the diagram above what percent of the area bounded by x<=5

Answer: The height of the first half of the curve is given by the formula y=.04x (You should verify this by the techniques of algebra.) So the height at x=5 is y=.20. The area forms a triangle whose area is A=.5*5*.20)= .5. The percent of the area is .5/1 = 50%.

5:1n Random Variables

That makes $x = 5$ the median of the distribution. Remember that the median is the "equal area" point of the curve. Is it also the mean? Since the function is symmetrical around 5, the answer is yes! Remember that the mean is the point upon which the curve would balance. In this case it is easy to see that 5 is that point. Later when we discuss more complicated functions it will not always be so easy. In those cases, we will just have to trust the technology and the tables.