RANDOM VARIABLES

Random Processes: A random process may be thought of as a process where the outcome is probabilistic (also called *stochastic*) rather than deterministic in nature; that is, where there is uncertainty as to the result.

Examples:

1. Tossing a die – we don't know in advance what number will come up.
2. Flipping a coin – if you carefully enough devise an apparatus to flip the coin, it will always come up the same way. However, normal flipping by a human being can be considered a random process.
3. Shaking up a collection of balls in a hat and then pulling out one without looking.

Caution: All the examples above may appear to be situations where the outcomes have equal probabilities. But consider

1. A die that is not fair – e.g., 2 comes up twice as often as 3
2. A coin that is not fair – e.g., heads comes up twice as often as tails
3. A collection of balls not all of the same size or weight – you are more likely to pick out large balls than small ones, or light ones than heavy ones.

Random Variables: In most applications, a ***random variable*** can be thought of as *a variable that depends on a random process*.

Examples:

1. Toss a die and look at what number is on the side that lands up.
    - Tossing the die is an example of a random process;
    - The number on top is the value of the random variable.

2. Toss two dice and take the *sum* of the numbers that land up.
    - Tossing the dice is the random process;
    - The sum is the value of the random variable.

3. Toss two dice and take the *product* of the numbers that land up.
    - Tossing the dice is the random process;
    - The product is the value random variable.

Examples 2 and 3 together illustrate: *The same random process can be involved in two different random variables*.

4. Randomly pick (in a way that gives each student an equal chance of being chosen) a UT student and measure their height.
    - Randomly picking the student is the random process.
    - The student's height is the value of the random variable.

5. Randomly pick (in a way that gives each student an equal chance of being chosen) a student *in a particular UT class* and measure their height.

- Picking the student is the random process.
- The student's height is the value of the random variable.

Examples 4 and 5 illustrate: *Using the same variable* (in this case, height) *but different random processes* (in this case, choosing from different populations) *gives different random variables*.

Confusing two random variables with the same variable but different random processes is a common mistake.

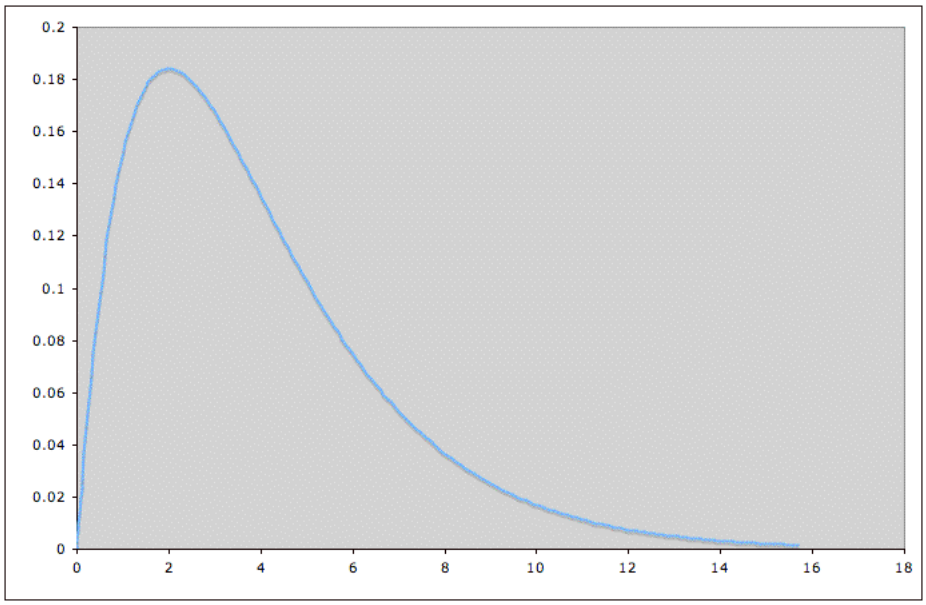
6. Measure the height of the third student who walks into the class in Example 5.
   - In all the examples before this one, the random process was done deliberately.
   - In Example 6, the random process is one that occurs naturally.
   - *Because Examples 5 and 6 depend on different random processes, they are different random variables*.

7. Toss a coin and see whether it comes up heads or tails.
   - Tossing the coin is the random process.
   - The value is heads or tails.
   - Example 7 shows that *a random variable doesn't necessarily have to take on numerical values*.

Probability Distributions: For random variables, probability enters as a *probability distribution*:

- Typically, some values (or ranges of values) of a random variable occur more frequently than others.
- For example, if we are talking about heights of university students, heights of around 5' 7" are much more common that heights of around 4' or heights around 7'.
- In other words, some values of the random variable occur with higher probability than others.
- This can be represented graphically by the ***probability distribution*** of the random variable.

Example:



- The possible values for the random variable are along the horizontal axis.
- The height of the curve above a possible value roughly tells how likely the nearby values are.
- This particular distribution tells us that values of the random variable around 2 (where the curve is highest) are most common, and values greater than 2 become increasingly less common, with values greater than 14 (where the curve is lowest) very uncommon.
- More precisely, *the area under the curve between two values a and b is the probability that the random variable will take on values between a and b*.
- In this example, we can see that the value of the random variable is much more likely to lie between 2 and 4 (where the curve is high, hence has a lot of area under it) than between 12 and 14 (where the curve is low, and hence has little area under it).