MATH 3070 Introduction to Probability and Statistics
Lecture notes
Random Variables and Probability Distributions

Objectives:

1. Define a random variable
2. Understand the difference between a discrete and continuous random variable
3. Construct a probability distribution

Random Variables and Probability Distributions

The outcome of any trial (or experiment) can take on any of the possible values in the sample space. In an experiment where we flip two coins and record the results, we could have any of the following: HH, HT, TH, or TT. When we roll a single fair die the results could be any number between 1 and 6, inclusive. Since the values can change from any given trial to any other this makes the results a variable. And since we don't know what the value might be that makes it random. The "box" that holds this random value is called a **random variable**.

> A **random variable** is a variable whose values are determined by chance.

> A **random variable** is a function that associates a real number with each element in the sample space.

The value in this box depends on chance and can be any of the values possible. These values could be either integer (whole) values or values such as "heads" or "tails". In these cases we call these values **discrete** and further classify the variables as discrete random variables. If the values could be more precisely defined to a more granular degree (written in real, or floating point, notation, with a decimal point) we call them continuous random variables. For our purposes we will concentrate on the discrete random variables (**d.r.v.**) for now. Discrete random variables have values that can be counted. Some examples of d.r.v. are

- The number of cars through the GA 400 toll booth in one day.
- Passengers through Hartsfield-Jackson Airport during Thanksgiving.
- The score on the math portion of the SAT or ACT.

Each value in a random variable has a portion of the total probability assigned to it. We know the total has to be 1 when we sum all the probabilities and each share of the total probability is between zero and one. The assignment of probability to each value is referred to as the **probability distribution**. This can be represented as either a function with output values or as a table.

> **Probability distribution:** A listing of the possible values and corresponding probabilities of a discrete random variable; or a formula for the probabilities. (Weiss, p 291)

To construct a probability distribution first identify the values for the discrete random variable. Then compute the probabilities for each of those values. Then construct a table with the values for the d.r.v. and the corresponding probability. There are two rules for constructing a probability distribution (Bluman, p. 184)

1. The sum of the probabilities of all events in the sample space must equal 1.
2. The probability of each event in the sample space must be between or qual to 0 and 1. That is, $0 \leq P(X) \leq 1$.

For example, if X is the d.r.v. in an experiment involving rolling one fair die, the possible values for the d.r.v. are 1,2,3,4,5, and 6. Since the die is fair we know each value is equiprobable (has equal probability) so we can simply divide the total probability by six to compute the amount of probability to assign to each value ($\frac{1}{6}$). Constructing our table we get something like this:

roll of die X	1	2	3	4	5	6
probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Of course, the probability does not have to be evenly distributed. If we look at students in high school, grades 9 through 12, and treat the grade level as a random variable (X) then we could construct a probability distribution that looks like this

grade level X	9	10	11	12
probability	0.300	0.262	0.230	0.208

Each of these probabilities is between zero and one and the sum is $.300 + .262 + .230 + .208 = 1.0$. This is a valid probability distribution.