Random Variables

Motivating Ex.:
Assume that all mens basketball teams playing this season are equally strong. We are interested in

$$Y = \# \text{ of points scored by NC State in a game.}$$

- Before each game, we know the population of possible values.
- Each value occurs with some probability.
- However, we do not know what will be the number of points scored by NC State during the next game.

The outcome is random, hence the # of points scored in a game is a **random variable**.

- A **Random Variable (RV)** is a real-valued function
  - Domain (values it takes in) = ______________________________
  - Range (values it outputs) = ______________________________

  An RV assigns a real number to each outcome in a sample space.

Two Types of RVs we'll discuss

- ______________________________: takes on finite or countably infinite # of values
- ______________________________: takes on a subset of intervals of real numbers

Why do we need to distinguish between these two types of RVs?

Basic Definition and Probability Distributions

Discrete Random Variables - An Example

- **Discrete random variable** assumes only a finite or countably infinite # of values
- Ex: Flip a coin 3 times - Let Y = # of heads from the 3 tosses
  - Range of Y?
  - Called ______________________________ of the RV
- Each outcome has a _________________________

  To describe the distribution, we need to describe the probability for each outcome in the support!

- Function $P(y) = P(Y = y)$ is called the ____
- Can be represented as a table:

Possible Values of Y	y_1	y_2	...	y_n
Probability for each value	$P(y_1) = P(Y = y_1)$	$P(y_2) = P(Y = y_2)$	...	$P(y_n) = P(Y = y_n)$

Let's find the probability distribution for Y = # of heads from 3 tosses using a table:

Some other examples of discrete random variables:

- Y = # of textbooks purchased in a semester. Support:

- X = # of plants that bloom from a group of 20 plants. Support:

- Y = # of flips of a coin before first head. Support:

Probability distribution for a discrete random variable must follow the following rules:

- For every y in the support of the RV Y,

- The sum of the probabilities over the entire support must be 1.

Let's check for the coin example.

Rules of probability still apply. For any two distinct values in the support, call them y_1 and y_2

Let's compute $P(Y \geq 2)$ for the coin example.

Summary Characteristics of RVs

Just as in the numerical summaries section we will want to summarize characteristics of the distribution. What are the two major characteristics?

To find the ____ of a **discrete RV**

Let's find the mean of Y from the coin example.

To find the _______________________________ of a **discrete RV**

Let's find the variance of Y from the coin example.

Let X (Any capital can be used to denote a RV) denote the # of male children if a family has 2 children (assume a the probability of a male child is 0.4 and that the children are independent).

- Determine the support of X
- Find $P(x)$, the probability distribution of X using a table
- Show that $P(x)$ meets the two conditions to be a probability distribution for a discrete RV.
- Find $P(X = 0 \ or \ X = 2)$
- Find the average number of male children.
- Find the variance of the number of male children.

Binomial Distribution

Recognizing a Distribution

- Note: # of Heads example and the # of male children example are similar.
- Similar experiments with similarly defined RV's yield the same ________________________________
- This particular distribution so common, it is called the ________________________________
- Knowing and being able to recognize common distributions will save us from having to derive things over and over!

When does a RV follow the Binomial Distr.?

Consider the following experiments:

- a coin is flipped, the outcome is either a head or a tail.
- a baby is born, the baby is either born in March or is not.

In each of these examples, an event has two ___________________________________.

For convenience, one of the outcomes can be labeled __________________________ and the other outcome ________________________________

Bernoulli Trials

- An experiment with only two possible mutually exclusive outcomes (such as S or F) is called a Bernoulli Trial
  - Bernoulli trials are the basis of three 'families' of distributions:
    * ________________________________ distribution
    * ________________________________ distribution
    * ________________________________ distribution
- For a trial denote the probability of success as
- Then the probability of failure is

We have a **Binomial Experiment** if:

1. Full experiment consists of a sequence of ________________________________

2. ____ on each trial (Bernoulli Trials)

3. Probability of success $P(S) = \pi$ is ________________________________, where $0 \leq \pi \leq 1$

Define the RV $Y =$ ____

Then Y is said to follow a binomial distribution.

We write ____ for convenience.

For $Y = \#$ of heads in three tosses

For $X = \#$ of male children from the two

(see example 4.5 and example 4.6 on pages 159/160 for practice picking out binomial experiments)

General form of the Probability Distribution for a Binomial RV

Ex: Suppose we have a Binomial Experiment with $n = 3$ trials and $P(S) = \pi$ where π is an unknown parameter

- Let Y be the # of successes – ____

Outcome	$P(\text{Outcome})$	y	Reps	$P(Y = y)$
SSS	$\pi\pi\pi = \pi^3$	3	1	π^3
SSF SFS FSS				
SFF FSF FFS				
FFF				

For general n, the event $Y = y$ occurs when there are exactly

- Consider one such outcome w/1^{st} y trials successful last $n - y$ failures:

$$SSS \cdots SFFF \cdots F$$

- Probability of this outcome?

- How many different sequences with exactly y successes in n trials?

The Probability distribution for $Y \sim Bin(n, \pi)$ is:

$$y = 0, 1, 2, \cdots, n, \;\; 0 \leq \pi \leq 1$$

Binomial Distribution Example

Suppose 60% of NCSU students favor closed-book exams. A random sample (outcomes independent) of 5 NCSU students is drawn.

1. Define Success/Failure, n, π, and a RV Y that follows the Binomial distribution

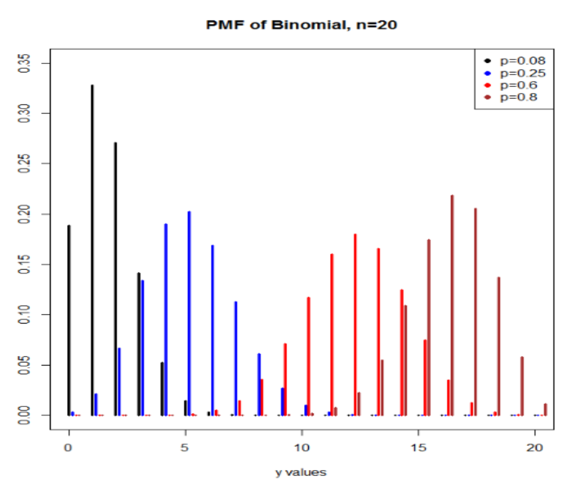
2. Calculate $P(exactly\ 1\ in\ favor)$

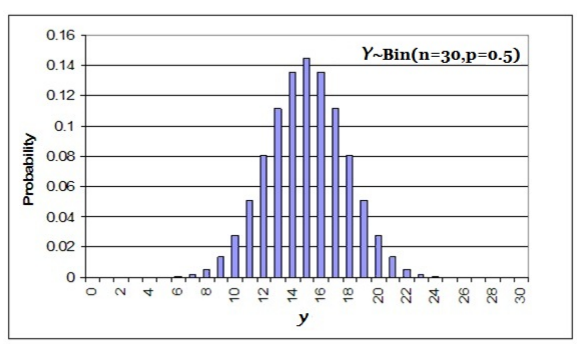
3. Calculate $P(less\ than\ 2\ in\ favor)$

4. Calculate $P(4\ or\ more\ in\ favor)$

(see examples 4.7 and 4.8 on page 162 for more practice with the binomial pmf)

We will want to have general formulas for the mean and variance of a binomial. Consider the following plots:





Binomial Expected Value - If $Y \sim Bin(n, \pi)$, then

Binomial Variance - If $Y \sim Bin(n, \pi)$, then

Binomial Standard Deviation =

Multiple Choice Test Example
Consider a multiple choice test with 20 questions, each with five possible answers (a,b,c,d,e), only one of which is correct. Let Y= # of questions guessed correctly.

1. Let's verify Y follows a binomial, calculate $E(Y)$, and calculate $Var(Y)$.

2. If scores of 50% and higher are passing, find the formula (i.e. don't simplify) for the probability of passing by guessing.

Many other common discrete distributions exist.

Connection with making inference

Hypothesis Testing Idea:

You love Pepsico and their products. They are having a promotion where their bottle caps are either winners (a free Pepsico product) or losers. Your friend claims you will hardly ever win, in fact he thinks only 1 in 20 bottles is a winner. You think the chance of winning is much higher than that.

To prove him wrong you grab 50 randomly selected Pepsico bottles and find that 12 of your caps are winners. How can we show your friend you are most likely correct?