Chapter 3: Discrete Distributions (Part 1)

STAT 5700: Probability

Prof. Katie Fitzgerald, PhD

Fall 2025

Contents

Discrete Random Variables & Their Probability Distributions (2.11, 3.1 - 3.2)

Remember our random babies example...

Probability = mathematical framework to describe and analyze random phenomenon

Random phenomenon = anything we cannot predict with certainty

E.g. randomly delivering 4 babies to 4 mothers

$S =$

1234 4	1243	1324	1342	1423	1432
2134 2	2143	2314	2341	2413	2431
3124	3142	3214	3241	3412	3421
4123	4132	4213	4231	4312	4321

Let Y = # of matches

Y is a **random variable**

A **random variable** is a **function** that maps each outcome in a sample space S (of a random experiment) to a real number.

$$Y : S \to \mathbb{R}$$

The **support** of Y is the set of real values that Y can take on.

$$\mathbb{S} = \{y : Y = y\}$$

We usually think of functions as deterministic, but in this case, the *input* to the function is random, so the resulting output is random as well.

In the random babies example, S was the set of 24 possible orderings of the 4 babies (e.g. 1234, 1243).

We defined Y to be the # of matches

Y is actually a **function**: e.g. $Y(1234) = 4$ and $Y(1243) = 2$

The **support** of Y is $\mathbb{S} = \{0, 1, 2, 4\}$. possible number of matches

We can think of a random variable as a numerical "summary" of some aspect of a random experiment

Why study random variables and the theory of probability? The probability of an observed event is often used to make inferences about a population. And events of interest are often numerical events that correspond to values of discrete random variables. For example:

- number of bacteria per unit area in the study of drug control on bacterial growth
- number of defective television sets in a shipment of 100 sets
- number of patients out of 10 that survive a disease
- amount of sales in USD
- When you collect data on a random sample of people from a population, each piece of information you collect (e.g., age, income, opinion on XYZ, transportation habits, etc) is considered a random variable. The way you randomly select people into your sample affects the probability of observing that sample, which in turn plays a major role in the inferences you are able to make about the whole population.

See Section 2.12 for a primer, but there is a whole branch of statistics called Survey Sampling that handles this in depth.

Example:
A coin is tossed three times, and the sample space is defined as

X = 3 2 2 1 2 1 1 0

$$S = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$$

Y = 1 1 0 0 1 1 0 0

T = 0 1 1 2 1 2 2 3

(1 = yes, heads on 2nd flip)
(0 = no, tails on 2nd)

Examples of random variables that are defined on S:

- X = the total number of heads
- Y = whether or not there is a heads on the 2nd flip
- Z = the number of heads minus the number of tails

$S_x = \{0, 1, 2, 3\}$

$S_y = \{0, 1\}$

$S_z = \{3, 1, -1, -3\}$

Define the **support** for each of the above random variables.

Example:
A rat is selected randomly from a cage and its sex is determined.

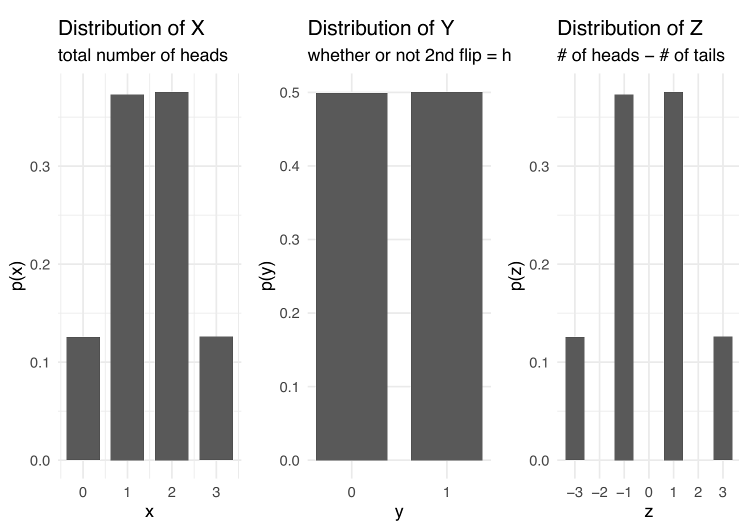
The sample space is therefore $S = \{\text{female, male}\} = \{F, M\}$.

Let Y be a function defined on S such that $Y(F) = 0$ and $Y(M) = 1$.

Y is a random variable, and its support is the set of numbers $\{y : y = 0, 1\}$

Distributions & random variables

We're often interested in the *behavior* of a random variable, how often it will take on the different values in its support. We describe this behavior using a **distribution**. Below is the distribution of the random variables Y, Y, Z defined above, based on 100,000 simulations of the random experiment of tossing 3 coins.



A key task in probability & statistics (and therefore in most of this class) is defining and working with mathematical models that describe the behavior of random variables. That is, we will be learning methods for defining and understanding the behavior of functions such as $p(y), p(y), p(z)$ above.

Because certain types of random variables occur so frequently in practice, it is useful to have at hand the probability for each value of a random variable (that is, to know its probability distribution). We will find that many experiments exhibit similar characteristics and generate random variables with the same type of probability distribution. Consequently, knowledge of the probability distributions for random variables associated with common types of experiments will eliminate the need for solving the same probability problems over and over again.

Common distributions we will study:

- **Bernoulli** - used to model success/failure events
  - Will a student graduate?
  - Will a basketball player make their free throw?
  - Is an email spam or not?
  - Does a sensor in a smart phone device detect motion or not?
- **Binomial** - used to model # of "successes" in a series of trials
  - Number of free throws made out of 10
  - Number of broken items in a shipment
  - Number of successful API requests out of 100 sent to a server.
  - Number of images correctly classified by a machine learning model out of 500 test cases
- **Geometric** - used to model # of "failures" until the first "success"
  - Number of people you poll until you find an independent voter
  - Number of items on a production line until the first defection item
  - How many joints are loaded (in welding) before the first beam fracture occurs
  - Number of retries needed to establish a reliable connection to a satellite.

Probability & Discrete random variables

Exercise:

A supervisor in a manufacturing plant has three men and three women working for him. He wants to choose two workers for a special job. Not wishing to show any biases in his selection, he decides to select the two workers at random. Let Y denote the number of women in his selection. Find the probability distribution for Y.

$S_Y = \{0, 1, 2\}$

$\binom{6}{2} = \frac{6!}{2!4!} = \frac{6 \cdot 5 \cdot 4!}{2 \cdot 4!} = 15$ ways to choose 2 workers

Y	$P(Y=y)$
0	$\binom{3}{0}\binom{3}{2} = 3/15$
1	$\binom{3}{1}\binom{3}{1} = 9/15$
2	$\binom{3}{2}\binom{3}{0} = 3/15$

A random variable Y is **discrete** if it can take on at most a countable number of values.

For a discrete r.v. Y, the probability $P(Y = y)$ is often denoted by $p(y)$ and is called the **probability distribution of Y**.

In the above example, $P(Y = 2) = p(2) = 3/15$

Y = random variable

y = stand-in for the values the r.v can take

Theorem 3.1
For any discrete probability distribution, the following must be true:

(a) $0 \leq p(y) \leq 1, \quad$ for all $y \in \mathbb{S}$ b/c they are probabilities
(b) $\sum_{y \in \mathbb{S}} p(y) = 1$

Exercise: For the previous example, define $p(y)$ for each value of Y and verify it is a valid probability distribution

$p(0) = 3/15$
$p(1) = 9/15$
$p(2) = 3/15$
$15/15$ ✓

$\sum_{y=0}^{2} p(y) = p(0) + p(1) + p(2) = 1$ ✓

We usually assume $p(y) = 0$ when $y \notin \mathbb{S}$. Recall $\mathbb{S}$ is the **support** of Y since it is the set of all unique values Y can take on (with positive probability).

Exercise:
Define the probability distribution for a 6-sided die. $S = \{1, 2, 3, 4, 5, 6\}$

$p(y) = \frac{1}{6} \quad y = 1, 2, \ldots, 6$

What would the probability distribution be for the result of a four-sided die? A 20-sided die? Can you come up with a general formula for the probability distribution of the result of an m-sided die?

$p(y) = \frac{1}{4} \quad y = 1, 2, 3, 4$

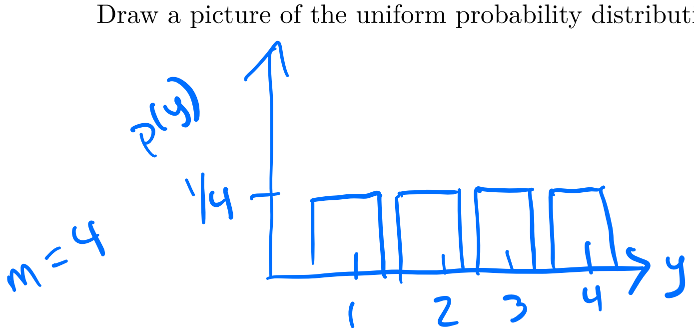
$p(y) = \frac{1}{20} \quad y = 1, 2, \ldots, 20$

$p(y) = \frac{1}{m} \quad y = 1, 2, \ldots m$

Uniform Distribution

In the previous example, we constructed the probability distribution for what's known as the **(discrete) uniform distribution**. In general, when a probability distribution is constant on the support of Y, we say that Y "follows a **uniform distribution**" (or sometimes "is **uniformly distributed**") $Y \sim U(a, b)$

Draw a picture of the uniform probability distribution:

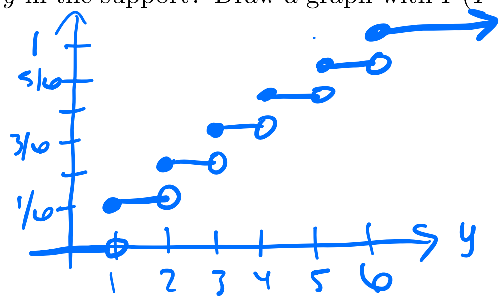


flat histogram
b/c prob. dist. is constant
(doesn't depend on y)

Example: $S = \{1, 2, 3, 4, 5, 6\}$ $P(y) = \frac{1}{6}$ $y = 1, 2, \ldots 6$

Let's return to our example where Y is the result of a single roll of a normal 6-sided die. What is $P(Y \leq y)$ for each value y in the support? Draw a graph with $P(Y \leq y)$ on the y-axis. *Hint: it will be a step function*

Y	$P(Y \leq y)$
1	1/6
2	2/6
3	3/6
4	4/6
5	5/6
6	6/6



Write $P(Y \leq y)$ as a piece-wise function: (*let k be an integer*)

$$P(Y \leq y) = \begin{cases} 0 & , \quad y < 1, \\ k/6 & , \quad k \leq y < k+1, \\ 1 & , \quad y \geq 6. \end{cases}$$

Cumulative Distribution Function (cdf)

In the previous example we found cumulative probabilities $P(Y \leq y)$. Cumulative probabilities are often of interest, so we have a special name and notation for this function:

Definition:

$$F(y) = P(Y \leq y), \quad -\infty < y < \infty$$

is called the **cumulative distribution function (cdf)** of a random variable Y.

CDF of Uniform Distribution

Let Y have a discrete uniform distribution over the first m positive integers. Then its probability distribution is

$$p(y) = \frac{1}{m} \quad y = 1, 2, \ldots, m,$$

and its cdf is

$$* \quad F(y) = P(Y \leq y) = \begin{cases} 0, & y < 1, \\ \frac{k}{m} & k \leq y < k+1, \\ 1, & m \leq y. \end{cases}$$

Note that the cdf $F(y)$ is a step function with a jump size of $1/m$ for $y = 1, 2, \ldots m$.

Non-uniform example

The uniform distribution is very useful, but there are many scenarios when random variables do NOT follow a uniform distribution. In other words, often $P(Y = y)$ will not be equal to a constant but will be different for different values of y in the support. In these cases, the probability distribution will be a function of y.

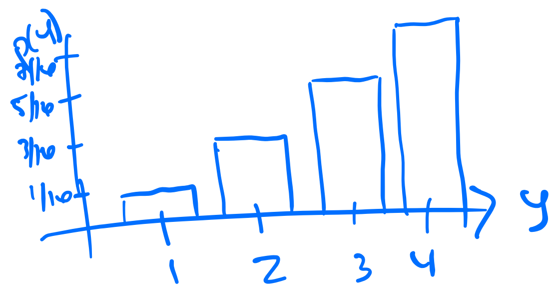
Example:
Roll a fair four-sided die twice, and let Y be the maximum of the two outcomes. The outcome space for this experiment is $S = \{(d_1, d_2) : d_1 = 1, 2, 3, 4; d_2 = 1, 2, 3, 4\}$, where we assume each of these points (d_1, d_2) has equal probability. What is the probability distribution of Y? *Hint: start by listing the possible outcomes* (d_1, d_2) and the corresponding values of Y

d_1 \ d_2	1	2	3	4
1	1	2	3	4
2	2	2	3	4
3	3	3	3	4
4	4	4	4	4

y	P(Y=y)
1	1/16
2	3/16
3	5/16
4	7/16

$$p(y) = \frac{2y-1}{16} \quad y = 1, 2, 3, 4$$

Draw a bar graph of the distribution of Y.



3.3 Expected Value

Exercise: A Game of Chance
Suppose you devise a game of chance that you're hoping you can convince your friends to play. The rules you come up with are that the participant rolls a fair 6-sided die; if they roll a 1,2, or 3 they win 1 dollar; if they roll a 4 or 5 they win 2 dollars, and if they roll a 6 they win 3 dollars. If the random variable Y represents the amount the participant is paid, the probability distribution of Y is given by:

$$p(y) = \frac{4-y}{6}, \qquad y = 1, 2, 3$$

How much should you charge your friends to play so that, on average, you break even? *Hint: think about the following three probabilities, and determine the average amount they will win per turn in the long run*

- $P(\text{they win } \$1)$ $= p(1) = \frac{4-1}{6} = \frac{3}{6}$
- $P(\text{they win } \$2)$ $= p(2) = \frac{4-2}{6} = \frac{2}{6}$
- $P(\text{they win } \$3)$ $= p(3) = \frac{4-3}{6} = \frac{1}{6}$

$\frac{3}{6}(1) + \frac{2}{6}(2) + \frac{1}{6}(3)$

$= \frac{10}{6} = \frac{5}{3} = 1\,2/3$

$= \$1.67$

Expected Value

Definition: If Y is a discrete random variable and $p(y)$ is its probability distribution, then the **expected value** of Y is defined as

$$E(Y) = \sum_{y \in \mathbb{S}} yp(y),$$

provided the summation exists.

In the previous example, $\mathbb{S} = 1, 2, 3$. Find $E(Y)$

$E(Y) = \sum_{y=1}^{3} y\left(\frac{4-y}{6}\right) = \cancel{\sum_{y=1}^{3}\left(\frac{4y-y^2}{6}\right)} = 1\left(\frac{3}{6}\right) + 2\left(\frac{2}{6}\right) + 3\left(\frac{1}{6}\right)$

$= \frac{5}{3} \Rightarrow \$1.67$

$E(Y)$ is often denoted by the Greek letter μ (pronounced "mu"), which is called the mean of Y or the mean of its distribution.

We can think of the expected value as the weighted mean, where the weights are the probabilities $p(y) = P(Y = y),\ y \in \mathbb{S}$.

Theorem 3.2: If Y is a discrete random variable with probability distribution $p(y)$ and support $\mathbb{S}$, then the expected value of a function $g(y)$ is given by

$$E[g(Y)] = \sum_{y \in \mathbb{S}} g(y)p(y),$$

provided the summation exists.

Example:
Suppose Y has the probability distribution $p(y) = 1/3$ for $S = \{-1, 0, 1\}$. Let $X = g(Y) = Y^2$. Find the expected value of X in two different ways:

1. First defining the probability distribution of X and then computing $E(X)$
2. Using the above theorem

①

Y	$X = Y^2$	$P(Y=y)$
-1	1	$1/3$
0	0	$1/3$
1	1	$1/3$

X	$P(X=x)$
0	$1/3$
1	$2/3$

$$E(X) = \sum_{x=0}^{1} x\,p(x) = 0(1/3) + 1(2/3) = \boxed{2/3}$$

②

$g(Y) = Y^2$

$$E(Y^2) = \sum_{y=-1}^{1} y^2 p(y) = (-1)^2(1/3) + \cancel{0^2(1/3)} + 1^2(1/3) = 2/3$$

Properties of Expected Value

$E(\cdot)$ is called a **linear operator**. For any random variable Y and constants a and b,

$$E(aY + b) = aE(Y) + b$$

PROOF:

$$E(aY+b) = \sum_{y \in S} (ay+b)\,p(y)$$

$$= \sum \left(ay\,p(y) + b\,p(y)\right)$$

$$= \sum_{y \in S} ay\,p(y) + \sum_{y \in S} b\,p(y)$$

$$= a \sum_{y \in S} y\,p(y) + b \underbrace{\sum_{y \in S} p(y)}_{=1}$$

$= 1$ b/c $p(y)$ is a prob dist.

$$= aE(Y) + b \quad //$$

Theorems 3.3 - 3.5 When it exists, the expectation $E(\cdot)$ satisfies the following properties:

1. If c is a constant, then $E(c) = c$.
2. If c is a constant, and g is a function, then
$$E[cg(Y)] = cE[g(Y)].$$
3. If c_1 and c_2 are constants and g_1 and g_2 are functions, then
$$E[c_1 g_1(Y) + c_2 g_2(Y)] = c_1 E[g_1(Y)] + c_2 E[g_2(Y)].$$

Example:
Let Y be a random variable with $E(Y) = 20$ and $E(Y^2) = 416$. Find

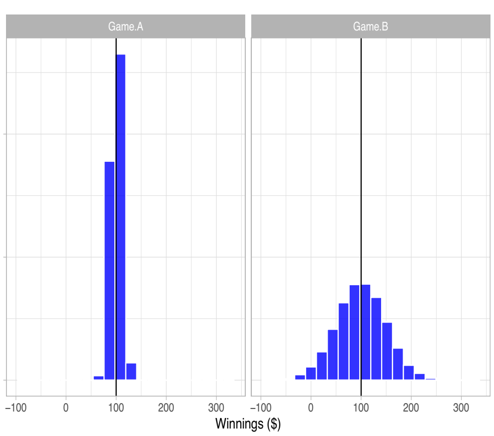
a. $E(4Y - 3)$
b. $E(2Y^2 - 6Y)$

a) $E(4Y-3) = E(4Y) - E(3) = 4E(Y) - 3 = 4(20) - 3 = \boxed{77}$

b) $E(2Y^2 - 6Y) = 2E(Y^2) - 6E(Y) = 2(416) - 6(20) = \boxed{712}$

expected value = mean = measure of center

Mean isn't the only thing that matters...



Special Expectations: Variance

Definition:
If Y is a random variable with $E(Y) = \mu$, then the **variance** of a random variable Y is defined to be the expected value of $(Y - \mu)^2$. That is,

$$E[(Y-\mu)^2] = \sum_{y \in \mathbb{S}} (y-\mu)^2 p(y) = \sigma^2$$

is called the variance of Y and is often denoted by $V(Y)$, $Var(Y)$, or σ^2 (pronounced "sigma squared").

Variance is a *measure of spread*: $(y - \mu)$ measures how far each value of y falls from the mean. We can think of variance as a *weighted sum* of squared distances from the mean. This is an example of a "sum of squares," which occur often and are very useful in statistics (e.g. regression, ANOVA).

Definition:
The positive square root of the variance is called the **standard deviation** and is denoted by the Greek letter σ.

It will often be convenient to use a shortcut formula when finding the variance.

Theorem 3.6 Let Y be a discrete random variable with probability density $p(y)$ and mean $E(Y) = \mu$. Then,

$$* \quad V(Y) = E(Y^2) - \mu^2 = E(Y^2) - (E(Y))^2$$

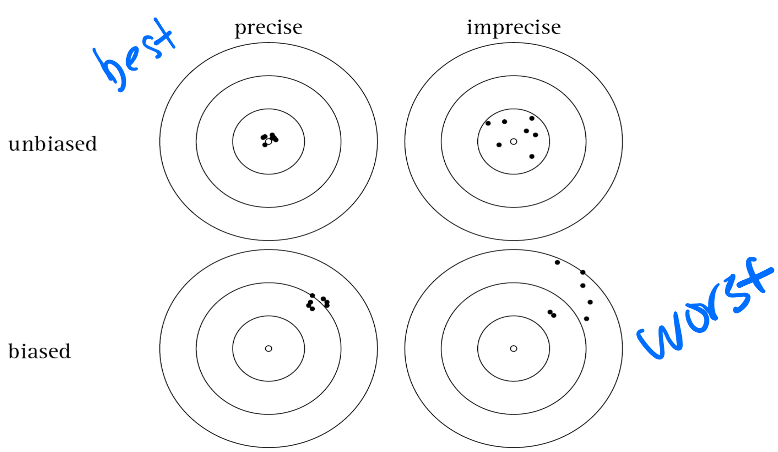
PROOF: $V(Y) = E[(Y-\mu)^2] = E[Y^2 - 2\mu Y + \mu^2]$ $\quad$ μ is a constant!

$= E(Y^2) - 2\mu E(Y) + E(\mu^2)$

$= E(Y^2) - 2\mu^2 + \mu^2$

$= E(Y^2) - \mu^2 = E(Y^2) - (E(Y))^2$

Statistical motivation for why we care about variance



Example:
Let Y have probability distribution $p(y) = 1/3$ for $y = -1, 0, 1$

1. Find the variance and standard deviation of Y

$$V(Y) = E(Y^2) - (E(Y))^2$$

$$E(Y) = 0 = \sum y\,p(y) = -1(1/3) + 0(1/3) + 1(1/3)$$

$$E(Y^2) = \sum_{y=-1}^{1} y^2 p(y) = (-1)^2(1/3) + 0^2(1/3) + 1^2(1/3) = 2/3$$

$$V(Y) = 2/3 - 0^2 = \boxed{2/3 = \sigma^2} \qquad \boxed{\sigma = \sqrt{\tfrac{2}{3}}}$$

2. Suppose the probability distribution is unchanged, but the support is instead $y = -2, 0, 2$. Note that this is a new random variable $X = 2Y$. How do the mean, variance, and standard deviation change? What do the variance and standard deviation measure?

$$E(X) = E(2Y) = 2E(Y) = 0$$

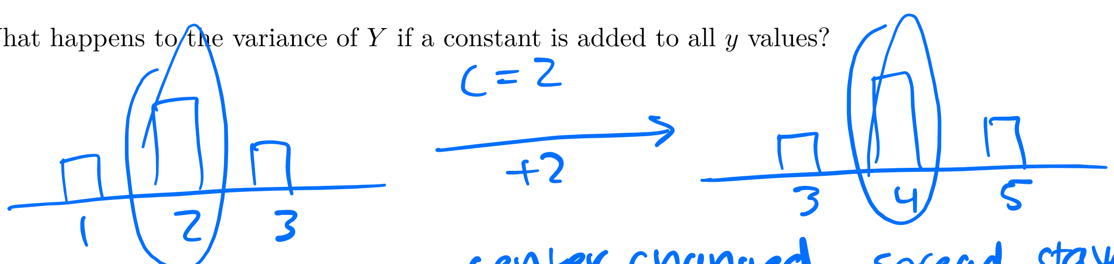
$$E(X^2) = E((2Y)^2) = 4E(Y^2) = 4(2/3) = 8/3$$

$$V(X) = E(X^2) - (E(X))^2 = 8/3 - 0^2 = 8/3 \qquad \sigma_X = \sqrt{8/3}$$

mean same   variance ↑   sd ↑

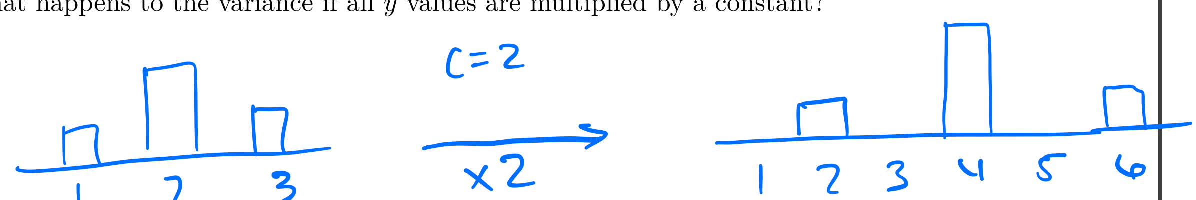
Properties of Variance

What happens to the variance of Y if a constant is added to all y values?



center changed, spread stayed same

What happens to the variance if all y values are multiplied by a constant?



both center and spread changed

↑ $c > 1$

↓ $c < 1$

$V(Y) = E(Y^2) - (E(Y))^2$

Theorem If Y is a random variable with finite variance, then for any constants a and b,

$$V(aY + b) = a^2 V(Y)$$

PROOF: $V(aY+b) = E[(aY+b)^2] - (E[aY+b])^2$

$= E[a^2Y^2 + 2aYb + b^2] - [aE(Y) + b]^2$

$= a^2E(Y^2) + \cancel{2abE(Y)} + \cancel{b^2} - [a^2(E(Y))^2 + \cancel{2abE(Y)} \cancel{+b^2}]$

$= a^2E(Y^2) - a^2(E(Y))^2$

$= a^2[\underline{E(Y^2) - (E(Y))^2}]$

$= a^2V(Y)$ //

Exercise:
Given $E(Y + 4) = 10$ and $E[(Y + 4)^2] = 116$, determine:

1. $\mu = E(Y)$
2. $V(Y + 4)$ *Hint: define* $X = Y + 4$ *and make use of the fact that* $V(X) = E(X^2) - (E(X))^2$
3. $\sigma^2 = V(Y)$

(1) $E(Y+4) = E(Y) + 4 = 10 \implies E(Y) = 6$

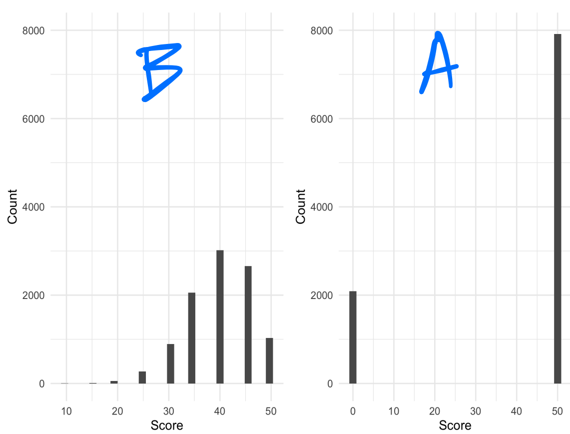
(2) $V(Y+4) = E((Y+4)^2) - (E(Y+4))^2$

$= 116 - 10^2 = 16$

(3) $V(Y+4) = V(Y) = 16$

Exercise - group activity
Suppose that Zane has a 20% chance of earning a score of 0 and an 80% chance of earning a score of 5 when he takes a quiz. Suppose also that Zane must choose between two options for calculating an overall quiz score: Option A is to take one quiz and multiply the score by 10, Option B is to take ten (independent) quizzes and add their scores.

1. Which option would you encourage Zane to take? Why? [No calculations yet!]
2. Which option do you suspect has a larger expected value, or do you suspect that the expected values will be the same? [Still no calculations! Just interested in your intuition]
3. Find the expected value and variance of his score on a single quiz, say Y_1.
4. Use properties of expected value to determine the expected value of his overall score with each option (A & B).
5. Which option do you suspect has a larger standard deviation, or do you suspect that the standard deviations will be the same? [Intuition only, no calculations]
6. Use properties of variance to determine the standard deviation of his overall score with each option. *Note: you can make use of the (new) fact that the variance of a sum of independent random variables is the sum of their individual variances. That is,* $V(Y_1 + Y_2 + \cdots + Y_{10}) = V(Y_1) + V(Y_1) + \cdots + V(Y_{10})$. If Zane's goal is to maximize his probability of obtaining an overall score of 50 points, which option should he select? Explain.
7. Calculate the probability, for each option, that Zane scores 50 points. Comment on how they compare.
8. The following graphs display the probability distributions of Zane's overall quiz score with these two options. Which graph goes with which option? Explain.



$V(Y) = E[(Y-\mu)^2]$

distance from the mean

3.4 Binomial Distribution

Bernoulli Experiments

A **Bernoulli experiment** is a random experiment that has two mutually exclusive and exhaustive outcomes, often referred to as "success" and "failure."

Lots of real world occurrences fit this binary structure (e.g. heads/tails, life/death, disease/no disease, graduate/not graduate). Note, sometimes choosing which of the two outcomes to call a "success" is arbitrary.

A sequence of **Bernoulli trials** occurs when a Bernoulli experiment is performed several *independent* times, and the probability of success, p, remains the same from trial to trial.

Example 1: flipping a fair coin 5 times corresponds to 5 Bernoulli trials with probability of success $p = 0.5$.
Example 2: Suppose the probability that a seed will germinate is 0.8, and we consider germination to be a "success." If we plant 10 seeds and can assume the germination of one seed is independent of the germination of another seed, this corresponds to 10 Bernoulli trials with $p = 0.8$.

Bernoulli Distribution

Let the random variable Y be the outcome of a Bernoulli experiment, where $Y = 1$ denotes a "success" and $Y = 0$ denotes a "failure." The probability distribution of Y is given by

$$p(y) = p^y(1-p)^{1-y}, \quad y = 0, 1,$$

$P(Y=0) = p(0) = p^0(1-p)^{1-0} = 1-p$

$P(Y=1) = p(1) = p^1(1-p)^{1-1} = p$

and we say that Y follows the **Bernoulli distribution**, or that Y is a **Bernoulli** random variable.

Sometimes we will denote the probability of failure as $q = (1-p)$.

Exercise: Find the mean and variance for the Bernoulli distribution. That is, find $E(Y)$ and $V(Y)$ when Y is a Bernoulli random variable.

$E(Y) = \sum_{y=0}^{1} y\,p(y) = 0(1-p) + 1(p) = \boxed{p}$

$V(Y) = E(Y^2) - (E(Y))^2$

$E(Y^2) = \sum_{y=0}^{1} y^2 p(y) = 0^2(1-p) + 1^2(p) = p$

$V(Y) = p - p^2 = \boxed{p(1-p)}$

Exercise:
Joel Embiid, star player drafted as the #3 overall pick by the Philadelphia 76ers, made 86% of his free throws in his rookie season. Suppose Embiid is fouled while attempting a 3 point shot with 1 second left in the game and the Sixers down by 2 points. What is the probability the Sixers win the game? Tie? Lose? Develop a random variable (and probability distribution) to answer these questions.

Y = # of makes

Y	$P(Y=y)$	
0	$\binom{3}{0}(.14)^3(.86)^0$	.0027
1	$\binom{3}{1}(.86)^1(.14)^2$	.0506
2	$\binom{3}{2}(.86)^2(.14)^1$	.311
3	$\binom{3}{3}(.86)^3(.14)^0$	.636

$p(y) = P(Y=y) = \binom{3}{y}(.86)^y(.14)^{3-y}$

$P(\text{win}) = 0.636$

$P(\text{tie}) = 0.311$

$P(\text{lose}) = 0.053$

Binomial distribution

In a sequence of Bernoulli trials, we are often interested in the total number of successes but not the actual order of their occurrences.

If we let the random variable Y equal the number of observed successes in n Bernoulli trials, then the possible values of Y are $0, 1, 2, ..., n$.

If y successes occur, where $y = 0, 1, 2, ..., n$, then $n - y$ failures occur. The number of ways of selecting y positions for the y successes in the n trials is

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

Since the trials are independent and since the probabilities of success and failure on each trial are, respectively, p and $q = 1 - p$, the probability of each of these ways is $p^y(1-p)^{n-y}$.

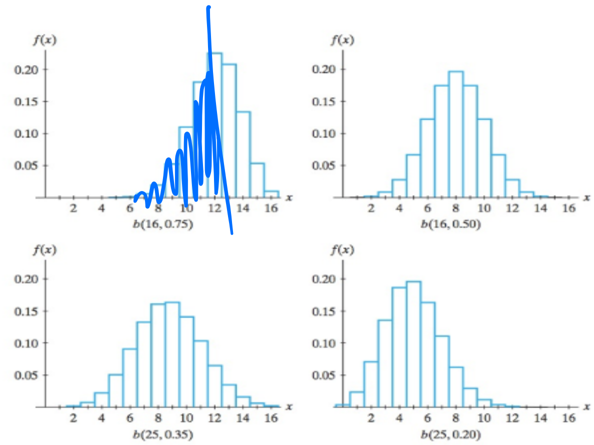
Thus, $p(y)$, the probability distribution of Y, is the sum of the probabilities of the $\binom{n}{y}$ mutually exclusive events; that is,

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

# of successes in n trials $\longrightarrow$ Binomial

If Y has probability distribution $p(y) = \binom{n}{y} p^y (1-p)^{n-y}$, we say Y follows the **binomial distribution** and the probabilities $p(y)$ are called the **binomial probabilities**.
A binomial experiment satisfies the following properties:

1. A Bernoulli (success–failure) experiment is performed n times, where n is a (non-random) constant.
2. The trials are independent.
3. The probability of success on each trial is a constant p; the probability of failure is $q = 1 - p$.
4. The random variable Y equals the number of successes in the n trials.



$Y \sim U(a,b)$

$Y \sim binom(n,p)$

Sometimes we write $Y \sim b(n,p)$, which is read as "Y is distributed as a binomial random variable with parameters n and p". That is, if we say $Y \sim b(23, 0.14)$, we mean Y is the number of successes in $n = 23$ Bernoulli trials that each have probability of success $p = 0.14$.

Exercise:
A manufacturing process has historically produced defective items every 1 out of 20. Five objects are selected independently from the production line. Y = # of defective items

$Y \sim b(n=5, p = 1/20 = .05)$

a. What is the probability none of the selected items are defective?

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

$p(0) = \cancel{\binom{5}{0}}\cancel{(.05)^0}(1-.05)^{5-0} = .95^5 = \boxed{.7738}$

b. What is the probability of 1 or fewer defective?

$P(Y \le 1) = P(Y=0) + P(Y=1) = p(0) + p(1) = \boxed{.9774}$

$p(1) = \binom{5}{1}(.05)^1(.95)^4 = .2036$

c. What is the probability more than 2 are defective?

$\boxed{0.0012}$

$P(Y>2) = 1 - P(Y \le 2)$

$p(2) -$

Binomial cdf cumulative dist. function

Like in the previous example, cumulative probabilities are often of interest when working with the Binomial distribution. Recall that cumulative probabilities are given by the cumulative distribution function

$$F(y) = \underline{P(Y \leq y)}, \qquad -\infty < y < \infty.$$

The cdf of the Binomial distribution is given by

$Y \sim \text{binomial}(n, p)$

$P(y) = \binom{n}{y} p^y (1-p)^{n-y}$

$$F(Y) = P(Y \leq y) = \sum_{y=0}^{y} \underline{\binom{n}{y} p^y (1-p)^{n-y}}$$

Binomial expansion

The binomial expansion is given by

$$(a+b)^n = \sum_{y=0}^{n} \binom{n}{y} a^y b^{n-y}$$

Exercise:
Use the binomial expansion to show that the probability distribution of the Binomial distribution is a valid probability distribution.

WTS: $\sum_{y=0}^{n} \binom{n}{y} p^y (1-p)^{n-y} = 1$

$a = p \quad b = 1-p$

by binomial expansion

$\sum_{y=0}^{n} \binom{n}{y} p^y (1-p)^{n-y} = (p + 1 - p)^n$

$= (1)^n = 1$

$Y \sim \text{binomial}(n, p)$

$E(Y) = np$

$V(Y) = np(1-p)$

Chapter 3 Group Work

Problem 1

For each of the following, determine the constant c so that $p(y)$ satisfies the conditions of being a probability distribution for a random variable Y. That is, find c such that $\sum_{y \in S} p(y) = 1$.

a. $p(y) = y/c, \qquad y = 1, 2, 3, 4$
b. $p(y) = cy, \qquad y = 1, 2, 3, \ldots, 10$
c. $p(y) = c(1/4)^y, \qquad y = 1, 2, 3, \ldots$ *Hint: use the infinite series identity from calculus that tells us* $\sum_{n=1}^{\infty} a_1 (r)^{n-1} = \frac{a_1}{1-r}$

Problem 2

Recall the non-uniform example where we rolled a fair four-sided die twice and let Y be the maximum of the two outcomes. We determined that the probability distribution of Y was given by

$$p(y) = \frac{2y-1}{16}, \quad y = 1, 2, 3, 4.$$

Define the cdf of Y. That is define $F(y) = P(Y \leq y)$ for each value of y in the support.

Problem 3

(Should be done AFTER Simulation Activity) Return to our 70% shooter 3-free throw example. We found the probability distributions of the random variables X (total number of makes), Y (whether or not 2nd shot was a make), and Z (number of makes - number of misses), listed below. Use the probability distributions to find the expected value of each random variable.

$$p(x) = \begin{cases} 0.027, & x = 0 \\ 0.189, & x = 1 \\ 0.441, & x = 2 \\ 0.343, & x = 3 \end{cases}$$

$$p(y) = \begin{cases} 0.3, & y = 0 \\ 0.7, & y = 1 \end{cases}$$

$$p(z) = \begin{cases} 0.027, & z = -3 \\ 0.189, & z = -1 \\ 0.441, & z = 1 \\ 0.343, & z = 3 \end{cases}$$

Problem 4

Let $E(Y) = 4$. Find

a. $E(Y)$ when $Y = 2Y + 3$
b. $E(Z)$ when $Z = 7 - 5Y$
c. $E(32)$

Problem 5

Let $p(y) = \frac{y}{10}, \quad y = 1, 2, 3, 4.$ Find:

a. $E(Y)$
b. $E(Y^2)$
c. $E(Y(5 - Y))$

Problem 6

Let $E(Y) = 5$ and $V(Y) = 36$. Find:

a) $V(3Y + 7)$
b) $V(2 - Y)$
c) $E(Y^2)$
d) $E(5Y + 2Y^2)$

Problem 7

In a lab experiment involving inorganic syntheses of molecular precursors to organometallic ceramics, the final step of a five-step reaction involves the formation of a metal-metal bond. The probability of such a bond forming is $p = 0.2$. Let Y equal the number of successful reactions out of $n = 25$ such experiments.

a) What distribution is appropriate to model Y?
b) Write out the probability distribution of Y
c) Find the probability that $Y = 1$
d) Find the probability that Y is at least 1
e) Find the mean, variance, and standard deviation of Y

Problem 8

A random variable Y has a binomial distribution with mean 6 and variance 3.6. Find $P(Y = 4)$.

Problem 9

Return to the Joel Embiid example (he makes 87% of his free throws, he's fouled shooting a 3-pointer). Describe in detail how you could, in principle, perform a by hand simulation involving physical objects (e.g. coins, dice, spinners, etc.) to estimate the probability that the 76ers lose. Be sure to describe (1) what one repetition of the simulation entails, and (2) how you would use the results of many repetitions. Note: you do NOT need to compute any numerical values.

SIMULATION ACTIVITY

Let's return to our example of tossing three coins. Recall the sample space is:

$$S = \{hhh, hht, hth, htt, thh, tht, tth, ttt\}$$

We defined the following random variables:

- X = the total number of heads
- Y = whether or not there is a heads on the 2nd flip
- Z = the number of heads minus the number of tails

In this activity, we are going to consider this to be a **weighted coin with a 70% probability of landing on heads**. We could think of this like a basketball player with a 70% free throw percentage shooting three free throws.

Your job is to develop two things:

1) psuedocode for how to simulate this experiment and investigate the distributions of X, Y, and Z
2) the probability mass function (probability distribution) for X, Y, and Z (using mathematics & rules of probability)

Some tips are provided below to get you started.

Part 1 - psuedocode

Anytime we simulate a random experiment, it's important to ask ourselves a few questions (in order):

- How would I simulate one run of the random experiment?
- How would I calculate my random variable(s) of interest from one run of the experiment?
- How could I adapt my code to do this *many* times (e.g. 10,000+), and end up with *many* observed values of my random variable(s) of interest?
- How could I aggregate my *many* observations to summarize the distribution of my random variable(s) of interest?

Part 1A

Write psuedocode for the experiment described above for tossing three coins. *Hint: you need to create `toss1`, `toss2`, and `toss3`. The following code would generate one toss of a FAIR coin and save it into an object called `toss1` The first argument controls the number of runs of the experiment, and the second controls the probability of a "success". It will return the value `TRUE` for a "success" and the value `FALSE` for a "failure".*

```
toss1 <- rbernoulli(1, 0.5)
toss1
```

```
## [1] FALSE
```

Part 1B

Write psuedocode for creating the three random variables X, Y, and Z, based on your three tosses created in Part 1A.

Hint: R assigns the value 1 to `TRUE` and 0 to `FALSE`, and you can add and subtract these logical results. For example,:

```r
TRUE + TRUE
```

```
## [1] 2
```

```r
TRUE + FALSE
```

```
## [1] 1
```

Part 1C

How can you adapt your code from Parts A & B to do the experiment *many* times (e.g. 10,000 times)? You want to end up with three vectors X, Y, and Z, each with 10,000 random observations.

Note that arithmetic operators on vectors are computed element-wise in R.

```r
a <- c(1, 2, 3, 4, 5)
b <- c(6, 7, 8, 9, 10)
a + b
```

```
## [1]  7  9 11 13 15
```

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \\ a_4 + b_4 \\ a_5 + b_5 \end{bmatrix}$$

Part 1D

Describe in words how you could use a vector of 10,000 random observations to summarize the distribution of that random variable? For example, how might you estimate the following:

x	P(X = x) = p(x) =
0	P(X = 0) = p(0) =
1	P(X = 1) = p(1) =
2	P(X = 2) = p(2) =
3	P(X = 3) = p(3) =

Part 2: probability distributions

For each random variable, ask

- What is the support? (We did this in class)
- What are all the different ways its possible to end up with each value in the support?
- What is the probability of ending up with each value in the support? That is, for a random variable W, what is $P(W = w)$ for each value w in the support?

For example, there is only one way for $X = 0$: when the outcome is *ttt*. The probability of this happening is $(0.3)(0.3)(0.3) = 0.027$.

Use the above steps (three times) to produce a probability distribution for each random variable: X, Y, and Z.