PROBABILITY

WHAT IS PROBABILITY?

- Probability is a value between 0 and 1 that a certain event will occur
- For example, the probability that a fair coin will come up heads is 0.5
- Mathematically we write:

$$P(E_{heads}) = 0.5$$

WHAT ISPROBABILITY?

- In the above "heads" example, the act of flipping a coin is called a trial.
- Over very many trials, a fair coin should come up "heads" half of the time.

TRIALS HAVE NO MEMORY!

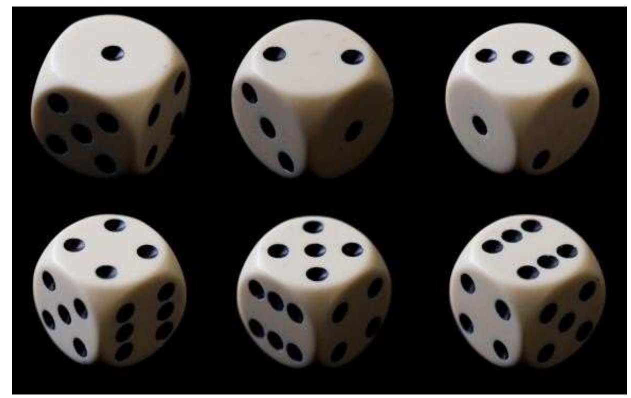
- If a fair coin comes up tails 5 times in a row, the chance it will come up heads is *still* 0.5
- You can't think of a series of independent events as needing to "catch up" to the expected probability.
- Each trial is independent of all others

EXPERIMENTS AND SAMPLE SPACE

- Each trial of flipping a coin can be called an experiment
- Each mutually exclusive outcome is called a simple event
- The sample space is the sum of every possible simple event

EXPERIMENTS AND SAMPLE SPACE

- Consider rolling a six-sided die
- One roll is an experiment
- The simple events are:



$E_1 = 1$ E2=2 E3=3

$E_4 = 4$ E5=5 E6=6

Therefore, the sample space is:

S = {E1 , E2 , E3 , E4, E5, E6}

EXPERIMENTS AND SAMPLE SPACE

- The probability that a fair die will roll a six:
- The simple event is:

E_6=6 (one event)

Total sample space:

S = $\{E_1, E_2, E_3, E_4, E_5, E_6\}$ (six possible outcomes)

The probability:

P(Roll Six) = 1/6

PROBABILITY EXERCISE

- A company made a total of 50 trumpet valves
- It is determined that one of the valves was defective
- If three valves go into one trumpet, what is the probability that a trumpet has a defective valve?

PROBABILITY EXERCISE

Calculate the probability of having a defective valve:

$$P(E_{defective\ valve}) = \frac{1}{50} = 0.02$$

PROBABILITY EXERCISE

Calculate the probability of having defective trumpet:

$$P(E_{defective\ trumpet}) = 3 \times P(E_{defective\ valve})$$
$$= 3 \times 0.02 = \mathbf{0.06}$$

PERMUTATIONS

- A permutation of a set of objects is an arrangement of the objects in a certain order.
- The possible permutations of letters a, b and c is:

abc	acb	bac	bca	cab	cba

PERMUTATIONS

- For simple examples like abc, we calculate the number of possible permutations as $n!$ *("n factorial")*
- abc = 3 items
- $n! = 3! = 3 \times 2 \times 1 = 6$ permutations

PERMUTATIONS

- You can also take a subset of items in a permutation
- The number of permutations of a set of n objects taken r at a time is given by the following formula:

$$_{n}P_{r} = \frac{n!}{(n-r)!}$$

PERMUTATIONS EXAMPLE #1

A website requires a 4 character password Characters can either be lowercase letters or the digits 0-9.

You may not repeat a letter or number.

How many different passwords can there be?

PERMUTATIONS SOLUTION#1

- Recognize that n, or the number of objects is 26 letters + 10 numbers = 36
- r, or the number of objects taken at one time is 4
- Plug those numbers into the formula:

$$_{36}P_4 = \frac{36!}{(36-4)!}$$

PERMUTATIONS SOLUTION#1

$$_{36}P_4 = \frac{36!}{(36-4)!} = \frac{36 \times 35 \times 34 \times 33 \times \cancel{32 \times 31 \ldots}}{\cancel{32 \times 31 \ldots}}$$

$$= 36 \times 35 \times 34 \times 33 = 1,\mathbf{413},\mathbf{720} \text{ permutations}$$

PERMUTATIONS ALLOWING REPETITION

- The number of arrangements of n objects taken r at a time, *with repetition* is given by

$$n^r$$

PERMUTATIONS EXAMPLE #2

How many 4 digit license plates can you make using the numbers 0 to 9 while allowing repetition?

PERMUTATIONS SOLUTION#2

Recognize there are 10 objects taken 4 at a time.
Plug that into the formula:

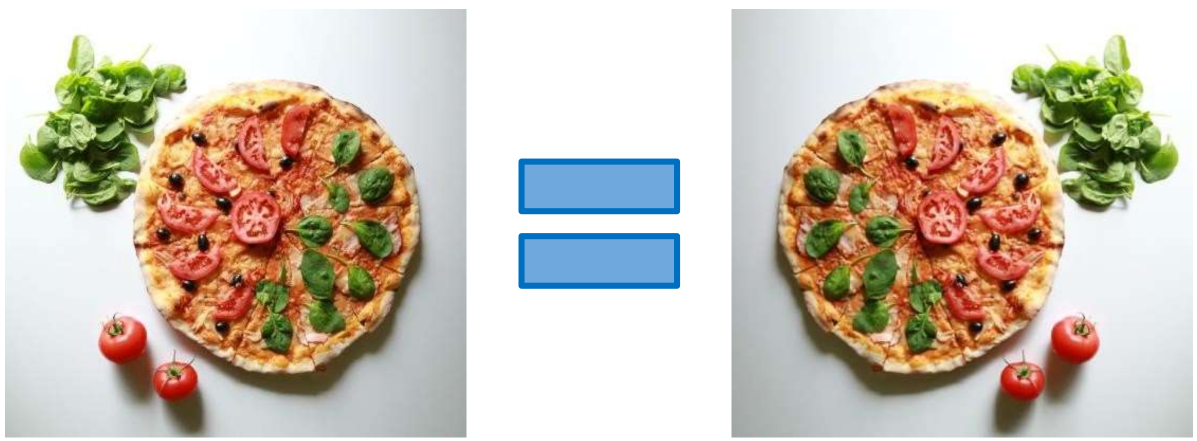
$$n^r = 10^4 = 10{,}000 \text{ permutations}$$

COMBINATIONS

- *Unordered* arrangements of objects are called combinations.
- A group of people selected for a team are the same group, no matter the order.

COMBINATIONS

- *Unordered* arrangements of objects are called combinations.
- A pizza that is half tomato, half spinach is the same as one half spinach, half tomato.



COMBINATIONS

- The number of combinations of a set of n objects taken r at a time is given by:

$$_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

COMBINATIONS VS. PERMUTATIONS

How many 3-letter combinations can be made from the letters ABCDE?

$${}_5P_3 = \frac{5!}{(5-3)!} = 5 \times 4 \times 3 = \mathbf{60}$$

ABC	ACB	BAC	BCA	CAB	CBA
ABD	ADB	BAD	BDA	DAB	DBA
ABE	AEB	BAE	BEA	EAB	EBA
ACD	ADC	CAD	CDA	DAC	DCA
ACE	AEC	CAE	CEA	EAC	ECA
ADE	AED	DAE	DEA	EAD	EDA
BCD	BDC	CBD	CDB	DBC	DCB
BCE	BEC	CBE	CEB	EBC	ECB
BDE	BED	DBE	DEB	EBD	EDB
CDE	CED	DCE	DEC	ECD	EDC

COMBINATIONS VS.PERMUTATIONS

How many 3-letter combinations can be made from the letters ABCDE?

Realize each row contains the same letters

ABC	ACB	BAC	BCA	CAB	CBA
ABD	ADB	BAD	BDA	DAB	DBA
ABE	AEB	BAE	BEA	EAB	EBA
ACD	ADC	CAD	CDA	DAC	DCA
ACE	AEC	CAE	CEA	EAC	ECA
ADE	AED	DAE	DEA	EAD	EDA
BCD	BDC	CBD	CDB	DBC	DCB
BCE	BEC	CBE	CEB	EBC	ECB
BDE	BED	DBE	DEB	EBD	EDB
CDE	CED	DCE	DEC	ECD	EDC

COMBINATIONS VS.PERMUTATIONS

How many 3-letter combinations can be made from the letters ABCDE?

Combinations:

$$C = \frac{n!}{r!(n-r)!} = \frac{5!}{3! \cdot 2!}$$

$$= \frac{5 \times 4 \times 3}{3 \times 2}$$

$$= \mathbf{10}$$

ABC	ACB	BAC	BCA	CAB	CBA
ABD	ADB	BAD	BDA	DAB	DBA
ABE	AEB	BAE	BEA	EAB	EBA
ACD	ADC	CAD	CDA	DAC	DCA
ACE	AEC	CAE	CEA	EAC	ECA
ADE	AED	DAE	DEA	EAD	EDA
BCD	BDC	CBD	CDB	DBC	DCB
BCE	BEC	CBE	CEB	EBC	ECB
BDE	BED	DBE	DEB	EBD	EDB
CDE	CED	DCE	DEC	ECD	EDC

COMBINATIONS EXAMPLE #1

For a study, 4 people are chosen at random from a group of 10 people.

How many ways can this be done?

COMBINATIONS SOLUTION#1

Since you're going to have the same group of people no matter the order they're chosen, you can set up the problem as a combination:

$$ {}_{n}C_{r} = \frac{n!}{r!(n-r)!} = \frac{10!}{4!(10-4)!} = 210 $$

COMBINATIONS EXAMPLE#1A

For a pizza, 4 ingredients are chosen from a total of 10 ingredients.

How many different combinations of pizza can we have?
In this situation we're only allowed to use each ingredient once.

COMBINATIONS SOLUTION#1A

Same as before, there will be 210 different types of pizza you can make:

$$ {}_{n}C_{r} = \frac{n!}{r!(n-r)!} = \frac{10!}{4!(10-4)!} = 210 $$

COMBINATIONS SOLUTION #1A

But what if we're allowed to repeat ingredients? (Use pepperoni 3 times and then add tomato once)

COMBINATIONS WITH REPETITION

- The number of combinations taken r at a time from a set n and allowing for repetition:

$$_{n+r-1}C_r = \frac{(n+r-1)!}{r!\,(n-1)!}$$

COMBINATIONS EXAMPLE#2

For a pizza, 4 ingredients are chosen at random from a possible of 10 ingredients.

How many different pizza topping combinations are there, allowing repetition?

COMBINATIONS SOLUTION#2

4 ingredients selected from 10 possible ingredients, allowing for repetition is:

$$ {}_{n+r-1}C_{r} = \frac{(n+r-1)!}{r!\,(n-1)!} = \frac{13!}{4!\,(9)!} = 715 $$

COMBINATIONS WITH/WITHOUTREPETITION

How many 3-letter combinations can be made from the letters ABCDE?

without repetition:

$${}_{n}C_{r} = \frac{n!}{r!(n-r)!} = \frac{5!}{3! \cdot 2!} = \mathbf{10}$$

ABC	ABD	ABE	ACD	ACE
ADE	BCD	BCE	BDE	CDE

with repetition:

$${}_{n+r-1}C_{r} = \frac{(n+r-1)!}{r!(n-1)!} = \frac{7!}{3!(4)!} = \mathbf{35}$$

ABC	ABD	ABE	ACD	ACE
ADE	BCD	BCE	BDE	CDE
AAA	AAB	AAC	AAD	AAE
BBA	BBB	BBC	BBD	BBE
CCA	CCB	CCC	CCD	CCE
DDA	DDB	DDC	DDD	DDE
EEA	EEB	EEC	EED	EEE

PERMUTATIONS & COMBINATIONS IN EXCEL

Order matters?	Repetition?	**In Excel**
Yes (permutation)	No	=PERMUT(n,r)
No (combination)	No	=COMBIN(n,r)
Yes (permutation)	Yes	=PERMUTATIONA(n,r)
No (combination)	Yes	=COMBINA(n,r)

PYTHON

How do we do it in python??

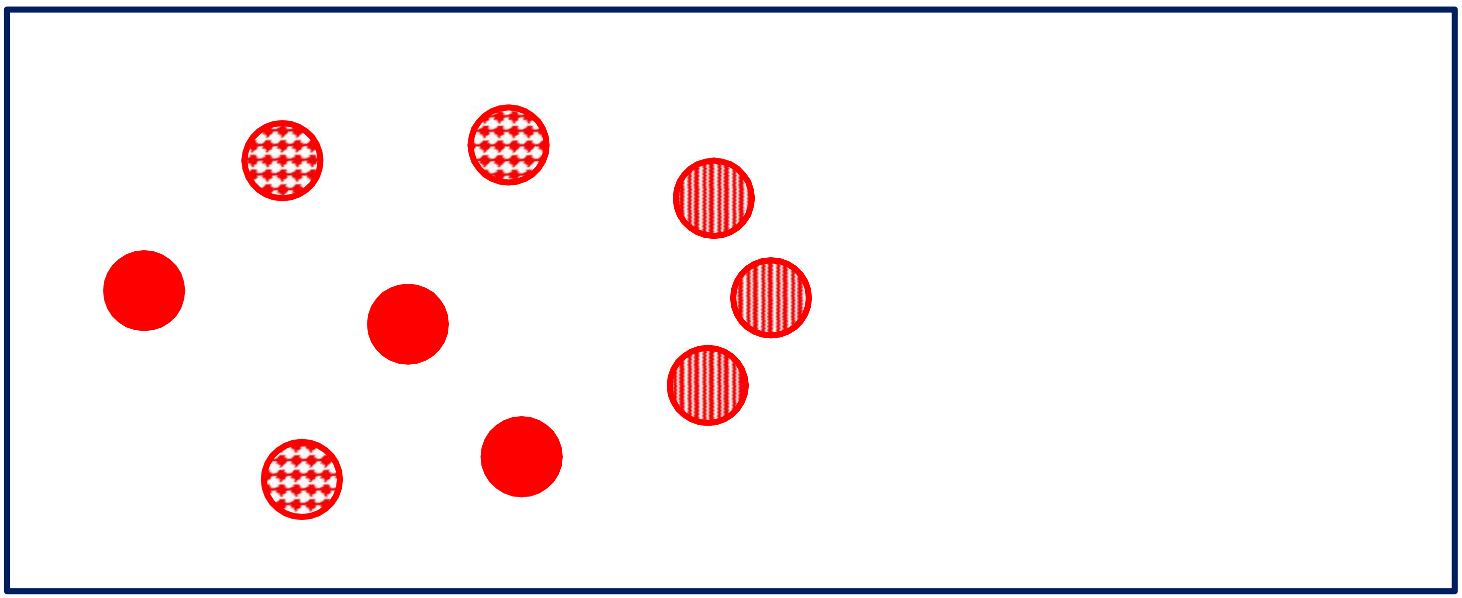
INTERSECTIONS, UNIONS AND COMPLEMENTS

INTERSECTIONS

- In probability, an intersection describes the sample space where two events *both* occur.
- Consider a box of patterned, colored balls

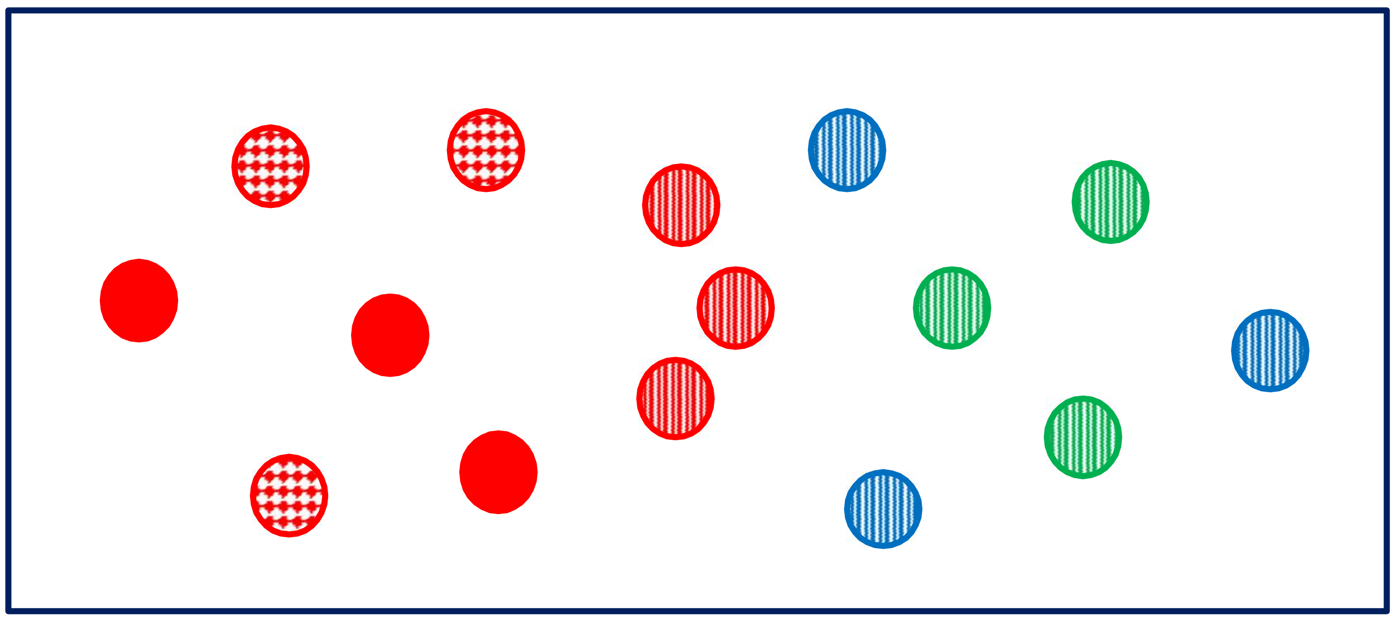
INTERSECTIONS

- 9 of the balls are red:



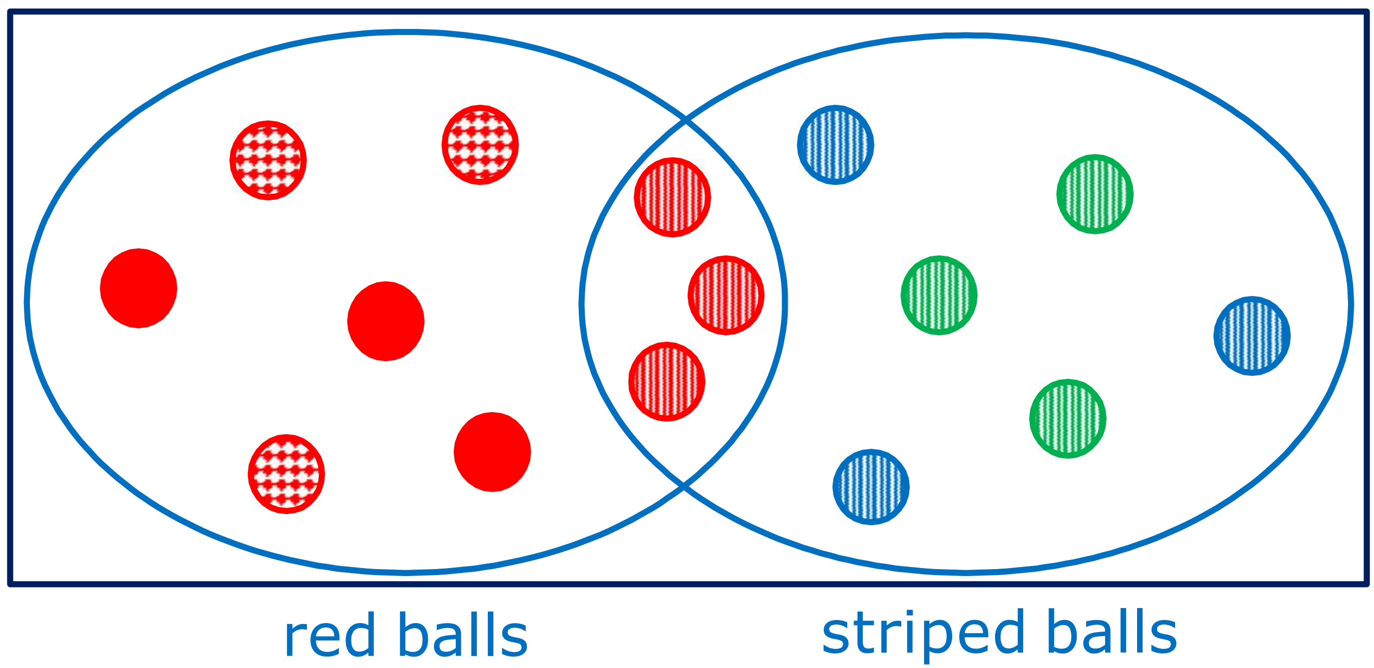
INTERSECTIONS

- 9 of the balls are striped:

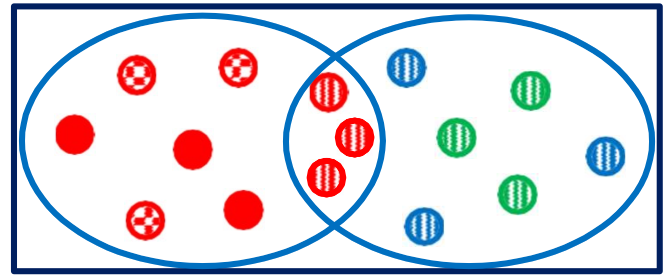


INTERSECTIONS

- 3 of the balls are both red and striped:



INTERSECTIONS



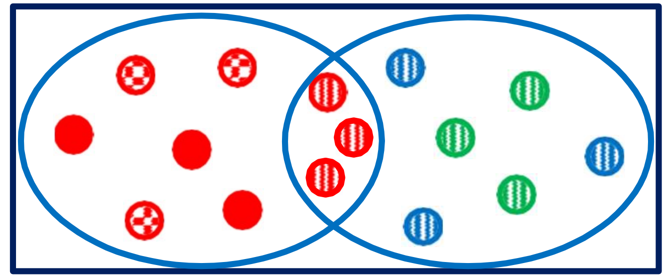
- If we assign A as the event of red balls, and B as the event of striped balls,
  the intersection of A *and* B is given as:

$$A \cap B$$

- Note that order doesn't matter:

$$A \cap B = B \cap A$$

INTERSECTIONS



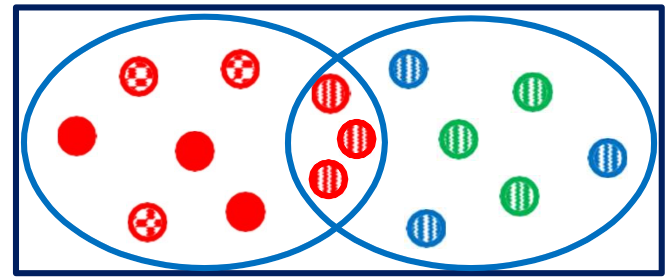
- The probability of A *and* B is given as

$$P\ (\ A\ \cap\ B\ )$$

- In this case:

$$P(A \cap B) = \frac{3}{15} = \mathbf{0.2}$$

UNIONS



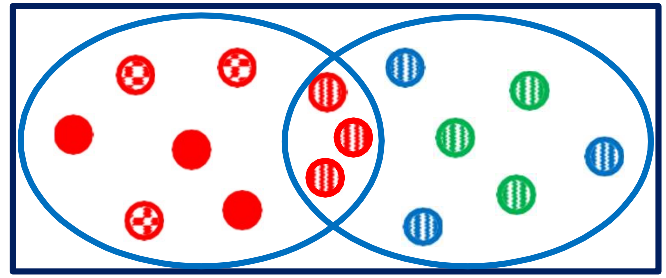
- The union of two events considers if A *or* B occurs, and is given as:

$$A \cup B$$

- Note again, order doesn't matter:

$$A \cup B = B \cup A$$

UNIONS



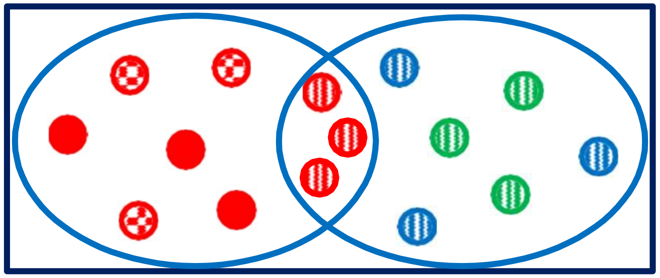
- The probability of A *or* B is given as:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- In this case:

$$P(A \cup B) = \frac{9}{15} + \frac{9}{15} - \frac{3}{15} = \frac{15}{15} = \mathbf{1.0}$$

COMPLEMENTS



- The complement of an event considers everything outside of the event, given by:

$$\bar{A}$$

- The probability of *not* A is:

$$P(\bar{A}) = 1 - P(A) = \frac{15}{15} - \frac{9}{15} = \frac{6}{15} = \mathbf{0.4}$$

INDEPENDENT AND DEPENDENT EVENTS

INDEPENDENT EVENTS

- An independent series of events occur when the outcome of one event has no effect on the outcome of another.
- An example is flipping a fair coin twice
- The chance of getting heads on the second toss is independent of the result of the first toss.

INDEPENDENT EVENTS

- The probability of seeing two heads with two flips of a fair coin is:

$$P(H_1 H_2) = P(H_1) \times P(H_2)$$
$$= \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

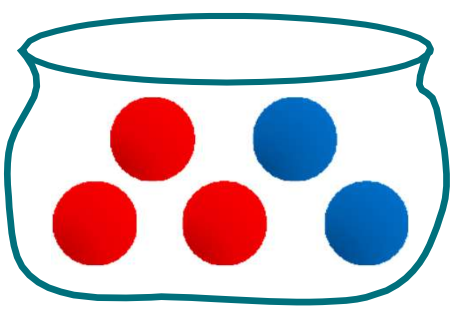
1st Toss	2nd Toss
H	H
H	T
T	H
T	T

DEPENDENT EVENTS

- A dependent event occurs when the outcome of a first event does affect the probability of a second event.
- A common example is to draw colored marbles from a bag *without replacement*.

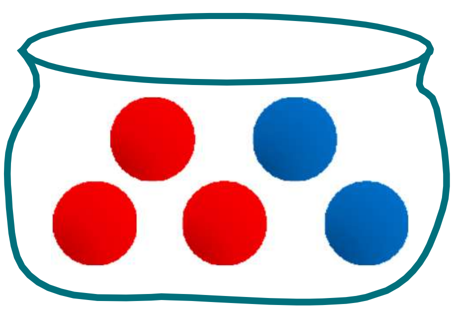
DEPENDENT EVENTS

- Imagine a bag contains 2 blue marbles and 3 red marbles.
- If you take two marbles out of the bag, what is the probability that they are both red?



DEPENDENT EVENTS

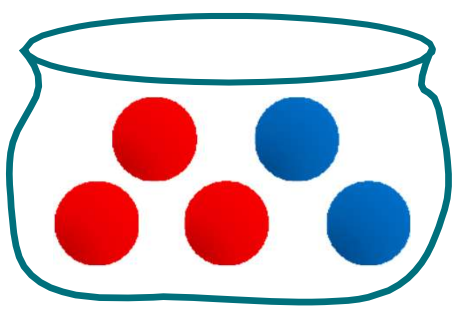
- Here the color of the first marble affects the probability of drawing a 2nd red marble.



DEPENDENT EVENTS

- The probability of drawing a first red marble is easy:

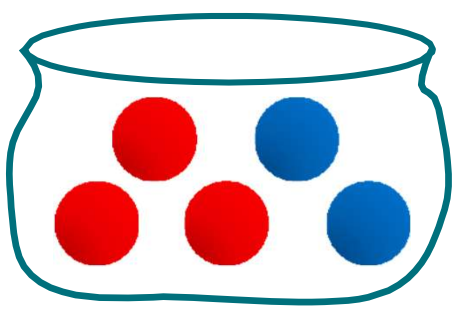
$$P(R_1) = \frac{3}{5}$$



DEPENDENT EVENTS

- The probability of drawing a second red marble *given that* the first marble was red is written as:

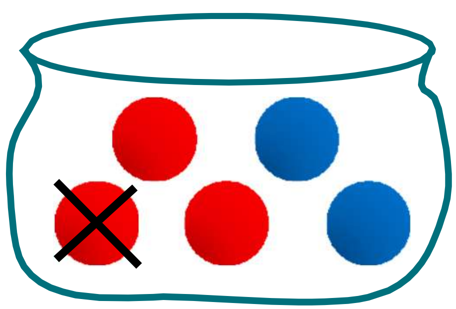
$$P(R_2|R_1)$$



DEPENDENT EVENTS

- After removing a red marble from the sample set thisbecomes:

$$P(R_2|R_1) = \frac{2}{4}$$

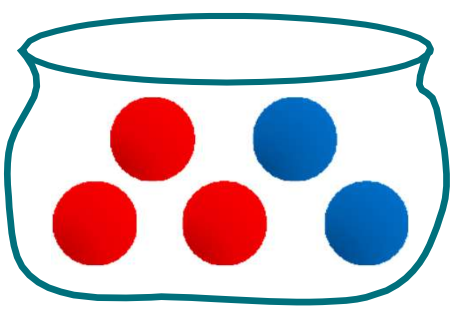


DEPENDENT EVENTS

- So the probability of two red marbles is:

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2|R_1)$$

$$= \frac{3}{5} \times \frac{2}{4} = \frac{6}{20} = \mathbf{0.3}$$



CONDITIONAL PROBABILITY

CONDITIONAL PROBABILITY

- The idea that we want to know the probability of event A, *given* that event B has occurred, is conditional probability.
- This is written as $P(A \mid B)$

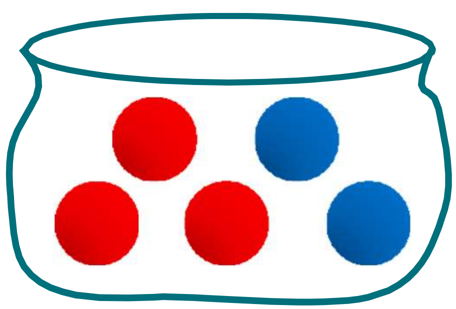
CONDITIONAL PROBABILITY

- Going back to dependent events, the probability of drawing two red marbles is:

$$P(R_1 \cap R_2) = P(R_1) \cdot P(R_2|R_1)$$

- The conditional in this equation is:

$$P(R_2|R_1)$$



CONDITIONAL PROBABILITY

- Rearranging the formula gives:

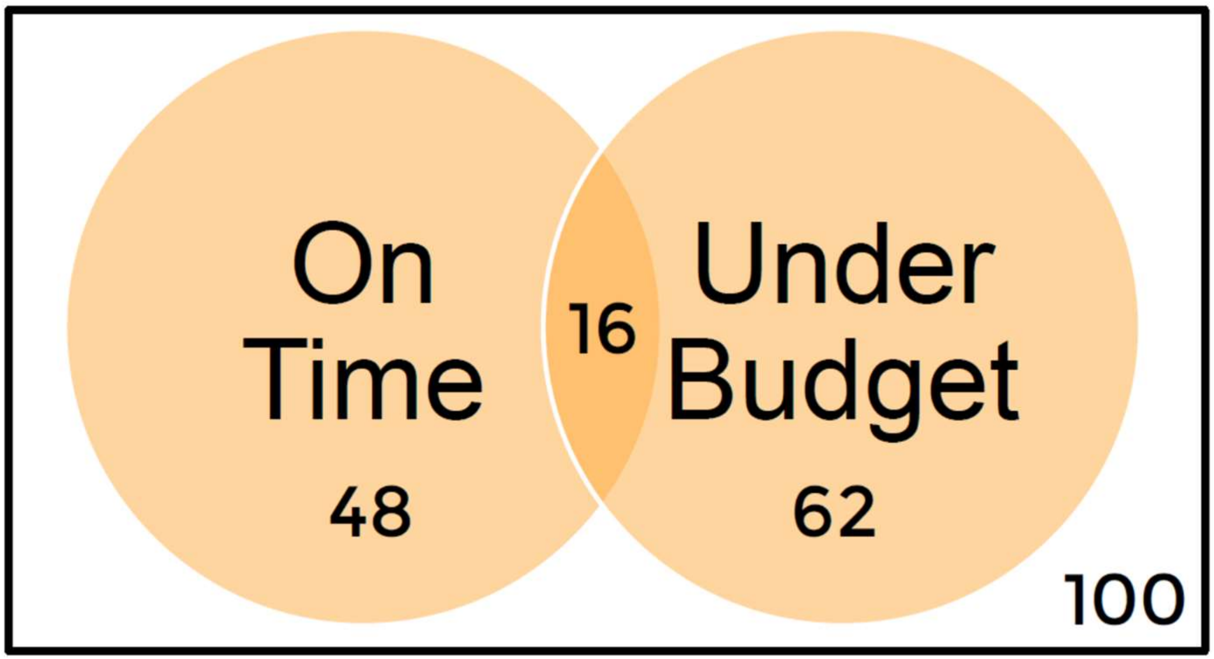
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- That is, the probability of A given B equals the probability of A and B divided by the probability of B

PRACTICAL SITUATION

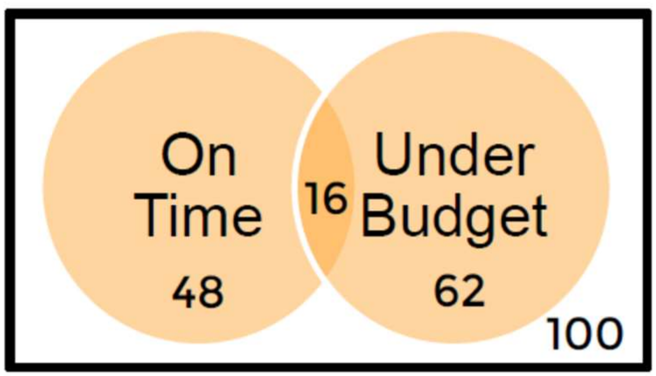
- A company finds that out of every 100 projects, 48 are completed on time, 62 are completed under budget, and 16are completed both on time and under budget.
- Given that a project is completed on time, what is the probability that it is under budget?

PRACTICAL SITUATION



PRACTICAL SITUATION

Given that a project is completed on time B, what is the probability that it is under budget A?

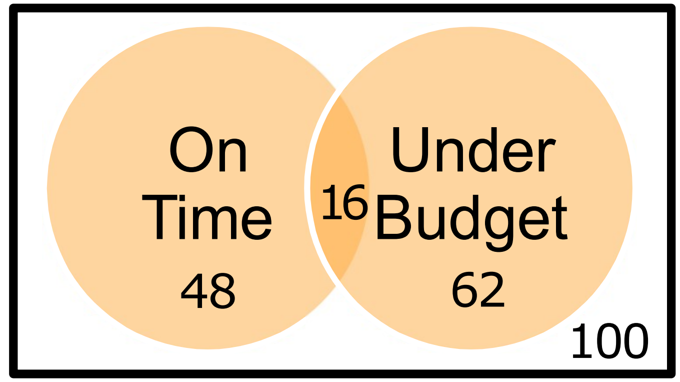


$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{16}{48} = \mathbf{0.33}$$

ADDITION RULE

- From our project example, what is the probability of a project completing on time *or* under budget?



- Recall from the section on unions:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- This is the addition rule

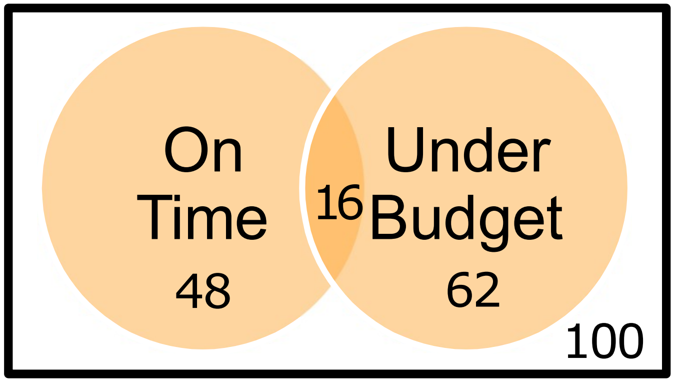
ADDITION RULE

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

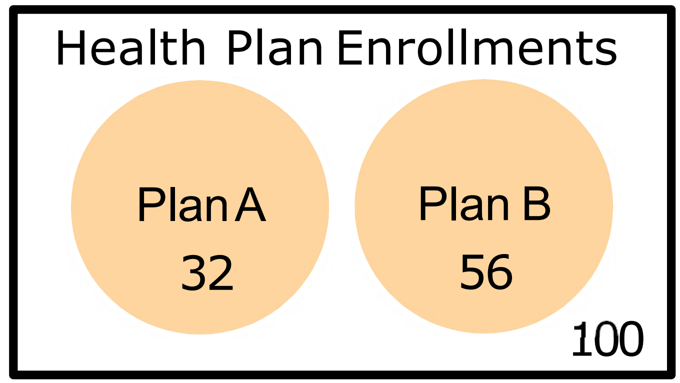
$$= \frac{48}{100} + \frac{62}{100} - \frac{16}{100}$$

$$= 0.48 + 0.62 - 0.16$$

$$= \mathbf{0.94}$$



ADDITION RULE FOR MUTUALLY EXCLUSIVE EVENTS



- When two events cannot both happen, they are said to be mutually exclusive.
- In this case, the addition rule becomes:

$$P(A \cup B) = P(A) + P(B) \cancel{- P(A \cap B)}$$

MULTIPLICATION RULE

- From the section on dependent events we saw that the probability of A and B is:

$$P(A \cap B) = P(A) \cdot P(B|A)$$

- This is the multiplication rule

MULTIPLICATION RULE
EXERCISE

- Given a standard deck of 52 cards, what is the probability of drawing 4 aces?

$$P(A \cap B \cap C \cap D) = P(A) \cdot P(B|A) \cdot P(C|AB) \cdot P(D|ABC)$$

$$= \frac{4}{52} \times \frac{3}{51} \times \frac{2}{50} \times \frac{1}{49} = \frac{24}{6,497,400} = \frac{1}{270,725}$$

BAYES THEOREM

BAYES THEOREM

- We've already seen conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \; provided\ that\ P(B) > 0$$

$$P(A \cap B) = P(A) \cdot P(B|A) \; provided\ that\ P(A) > 0$$

BAYES THEOREM

- We can then connect the two conditional probability formulas to get Bayes' Theorem:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)} \; provided\ that\ P(A), P(B) > 0$$

BAYES THEOREM

- Bayes Theorem is used to determine the probability of a *parameter*, given a certain event.
- The general formula is:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

PRACTICAL SITUATION

- A company learns that 1 out of 500 of their products are defective, or 0.2%.
- The company buys a diagnostic tool that correctly identifies a defective part 99% of the time.
- If a part is diagnosed as defective, what is the probability that it really *is* defective?

PRACTICAL SITUATION

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$P(A|B)$= probability of being defective if testing positive

$P(B|A)$= probability of testing positive if defective

$P(A)$= probability of being defective

$P(B)$= probability of testing positive

PRACTICAL SITUATION

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$P(B)$= probability of testing positive

= P(true positive) + P(false positive)

$$= P(B|A) \cdot P(A) + P(B|-A) \cdot P(-A)$$

$$P(B|-A) = 1 - P(B|A) = 1 - .99 = 0.01$$

$$P(-A) = 1 - P(A) = 1 - .002 = 0.998$$

PRACTICAL SITUATION

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|-A) \cdot P(-A)}$$

$$= \frac{0.99 \times 0.002}{0.99 \times 0.002 + 0.01 \times 0.998}$$

$$= 0.165$$

A positive test has a 16.5% chance of correctly identifying a defective part

PRACTICAL SITUATION

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|-A) \cdot P(-A)}$$

$$= \frac{0.99 \times \cancel{0.002}\ 0.165}{0.99 \times \underset{0.165}{\cancel{0.002}} + 0.01 \times \cancel{0.998}\ 0.835}$$

$$= \cancel{0.165}\ 0.951$$

What if we perform a second test, and that *also* comes up positive?

PRACTICAL SITUATION

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|-A) \cdot P(-A)}$$

$$= \frac{0.99 \times \cancel{0.002}\ 0.165}{0.99 \times \underset{0.165}{\cancel{0.002}} + 0.01 \times \cancel{0.998}\ 0.835}$$

$$= \cancel{0.165}\ 0.951$$

Two positive tests give us a 95.1% probability that the part is defective.

EXPECTED VALUES

Let X = number of prior convictions for prisoners at a state prison at which there are 500 prisoners. ($x = 0, 1, 2, 3, 4$)

$X = x$	0	1	2	3	4
$Number\ of\ Prisoners$	80	265	100	40	15
$f(x) = P(X = x)$	80/500	265/500	100/500	40/500	15/500
$f(x) = P(X = x)$	0.16	0.53	0.2	0.08	0.03

Answer

For this we need a weighted average since not all the outcomes have equal chance of happening (i.e. they are not equally weighted). So, we need to find our expected value of X, or mean of X, or $E(X) = \Sigma f(x_i)(x_i)$. When we write this out it follows:

$$= (0.16)(0) + (0.53)(1) + (0.2)(2) + (0.08)(3) + (0.03)(4) = 1.29$$

EXAMPLE

You buy one \$10 raffle ticket for a new car valued at \$15,000. Two thousand tickets are sold. What is the expected value of your gain?

	Win	Loose
X	14990	-10
P(X)	1/2000	1999/2000

HYPERGEOMETRIC DISTRIBUTION

In probability theory and statistics, the hypergeometric distribution is a discrete probability distribution that describes the probability of k successes (random draws for which the object drawn has a specified feature) in n draws, without replacement, from a finite population of size N that contains exactly K objects with that feature, wherein each draw is either a success or a failure.

In contrast, the binomial distribution describes the probability k successes in n draws with replacement.

HYPERGEOMETRIC DISTRIBUTION

Probability Mass Function	$P(X) = \dfrac{\dbinom{k}{x}\dbinom{N-k}{n-x}}{\dbinom{N}{n}}$
Mean	$\mu = E(X) = \dfrac{nk}{N}$
Variance	$\sigma^2 = V(X) = \left(\dfrac{N-n}{N-1}\right)(n)\left(\dfrac{k}{N}\right)\left(\dfrac{N-k}{N}\right)$

$N = total\ \#\ objects$

$n = \#\ sample$

$k = \#\ defective$

$x = \#\ successes$

EXAMPLE

Let's assume this box contains ten beans of the following flavours:

Banana
Blueberry
Booger
Earwax
Grass
Green Apple
Marshmallow
Rotten Egg
Lemon
Vomit

What is the chance that two yucky flavors are chosen?

How many beans are in the box?	$N = 10$
How many beans are chosen, at random, without replacement?	$n = 5$
How many yucky flavors are in the box?	$k = 4$
How many yucky flavors to be chosen?	$x = 2$

$$P(X) = \frac{\underbrace{\binom{k}{x}}_{\text{Yucky}}\underbrace{\binom{N-k}{n-x}}_{\text{Tasty}}}{\underbrace{\binom{N}{n}}_{\text{Total}}}$$

$$P(X) = \frac{\binom{4}{2}\binom{10-4}{5-2}}{\binom{10}{5}}$$

$$P(X) = \frac{\binom{4}{2}\binom{6}{3}}{\binom{10}{5}} = \frac{\left(\frac{4!}{2!(4-2)!}\right)\left(\frac{6!}{3!(6-3)!}\right)}{\left(\frac{10!}{5!(10-5)!}\right)} = \frac{(6)(20)}{252} = 0.476$$

BINOMIAL VS HYPERGEOMETRIC

Their differences lie in the way that sampling is done.

The binomial distribution has a fixed number of **independent trials**, whereas the hypergeometric distribution has a set number of **dependent trials**.

In other words, we will use the hypergeometric distribution whenever we have sampling without replacement!

Additionally, this distribution enables us to deal with situations arising when we sample from batches with a known number of defective items.

1

14000605