Extra material: https://onlinelibrary.wiley.com/doi/10.1111/test.12321

Module 2A : Probability

Frequentist and Bayesian building blocks

Agenda:

- **Frequentist probability**
  - Venn Diagram & definition of Event
  - Addition Rule
    - Mutually exclusive, not mutually exclusive
  - Multiplication Rule
    - Independent, not independent
  - Conditional Probability
  - Important differentiation: $P(A \cup B)$, $P(A \cap B)$, $P(A|B)$

Probability:

Two major splits in how probability is defined:

Frequency Interpretation:

Frequency of a particular outcome (an event) across many random trials

Subjective (Bayesian) Interpretation:

Subjective belief or opinion of the chance that a particular outcome (an event) will be realized

Random Experiment:

The process of observing the outcome of a chance event

ex - one roll of a die

Sample State Space:

A list of all possible elementary outcomes of an experiment

ex - {1, 2, 3, 4, 5, 6}

Event:

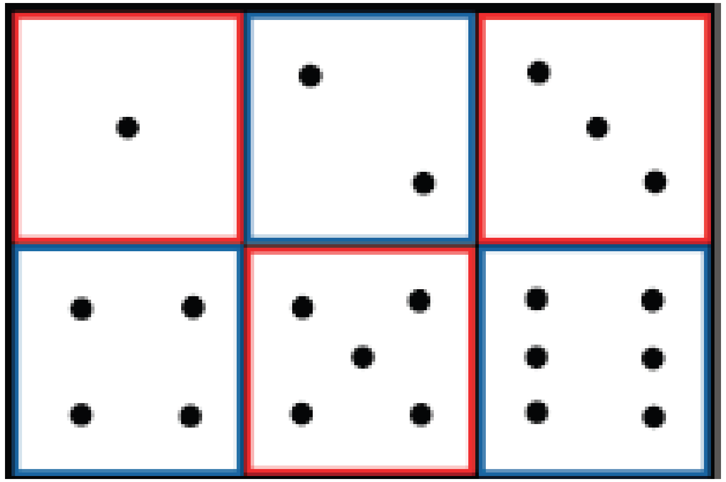
A set made up of elements from the sample space

ex - {1, 2, 3, 4, 5, 6}

ex - “an even number”

Sample Space examples:

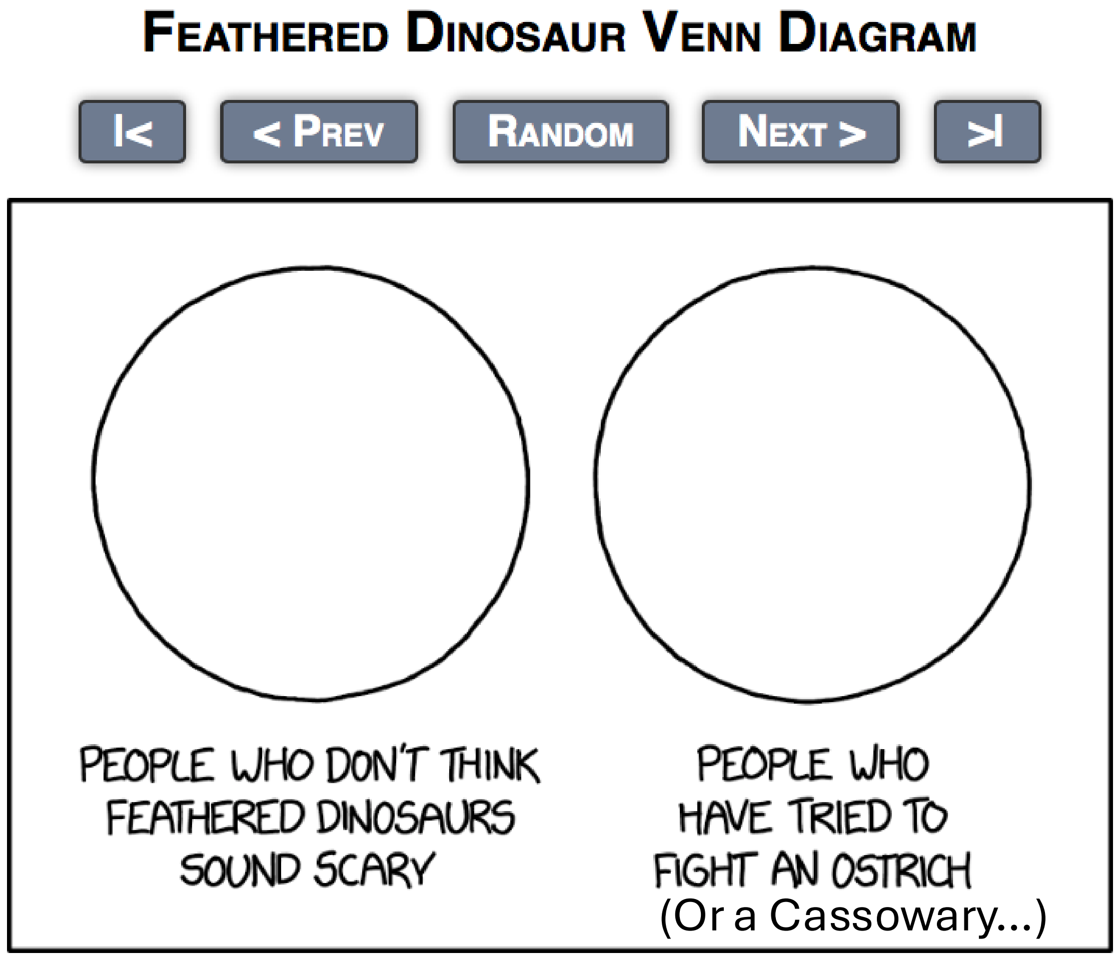
One Die Toss



{1,2,3,4,5,6}

One Coin Toss

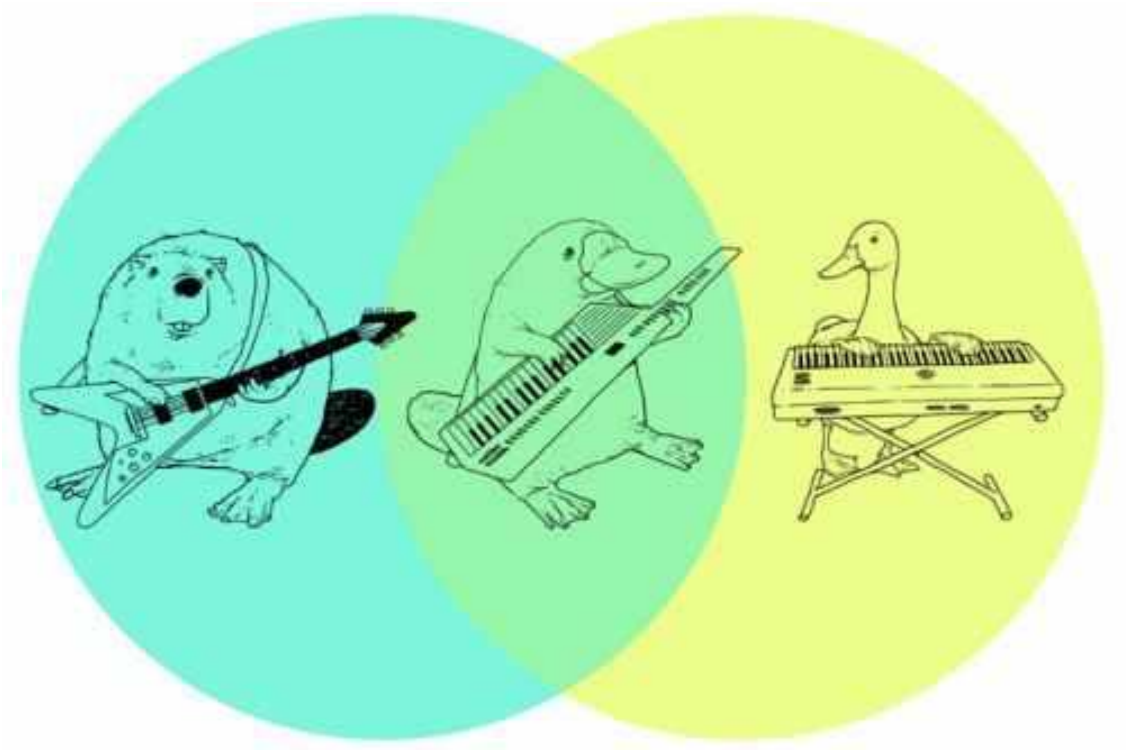
{heads, tails}



(Or a Cassowary...)

https://xkcd.com/2090/

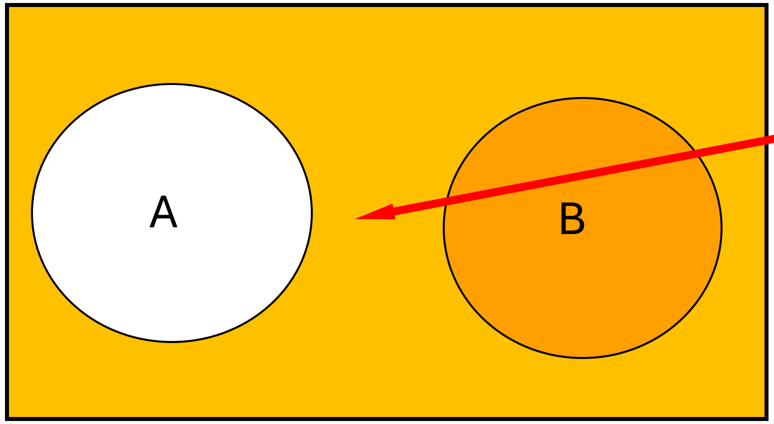
The BEST Venn diagram



http://i.imgur.com/CsaCFLc.jpg

NOT A

The event does not occur (complement); everything except A

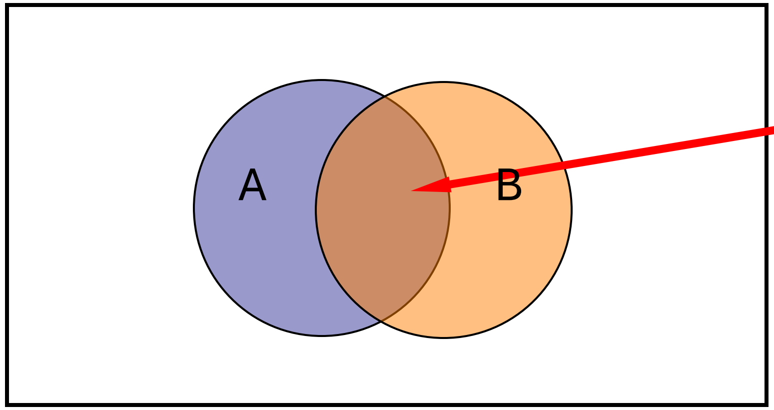


Mutually exclusive

$$P(A \cap B) = 0$$

AND

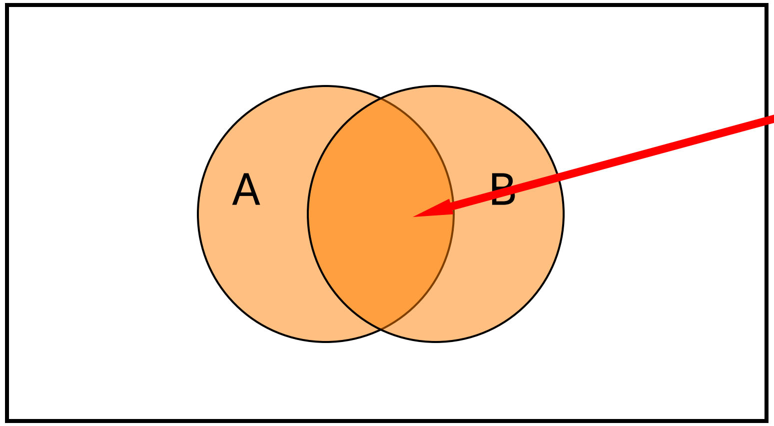
Both events occur (intersection)



$$P(A \cap B) \neq 0$$

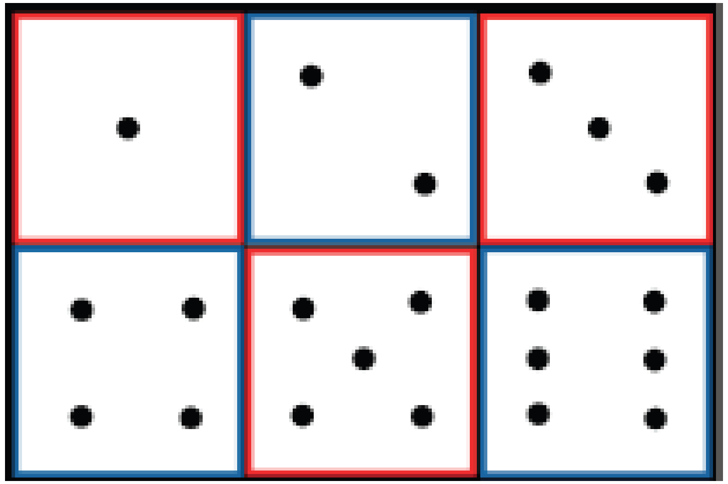
OR

Either events or both events occur (union)

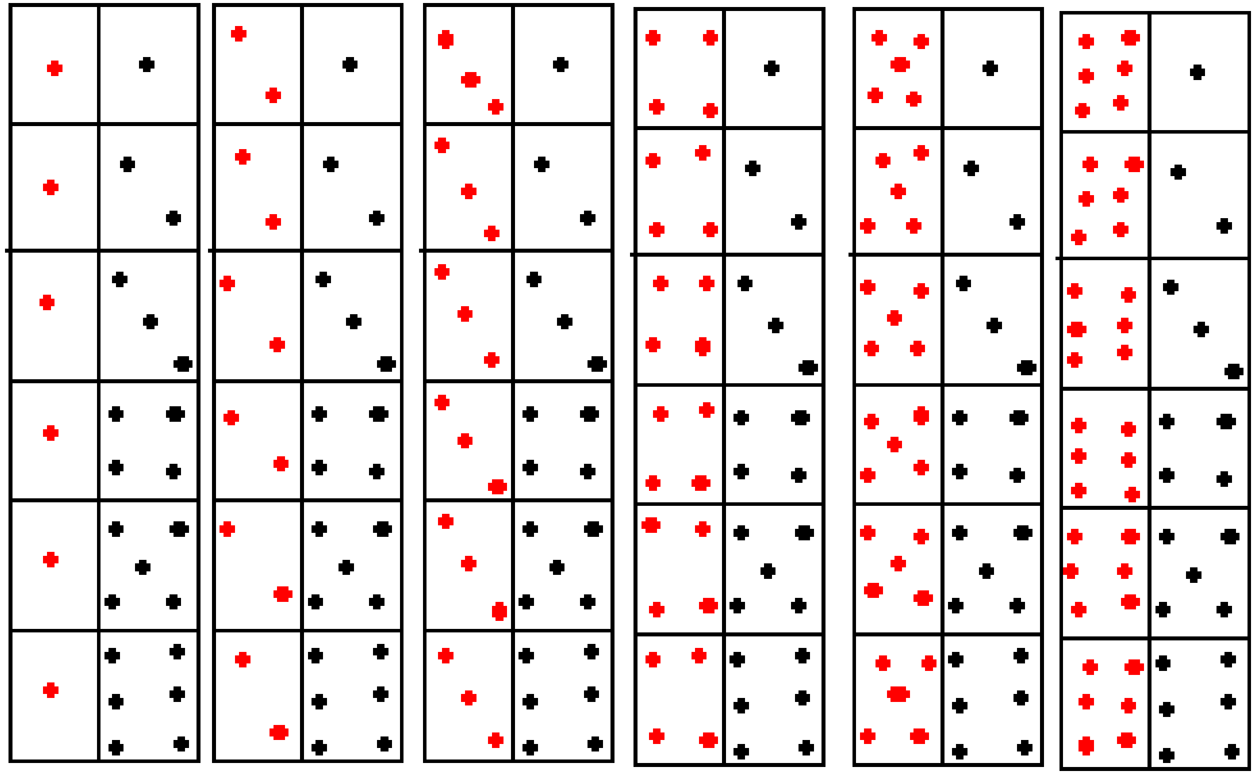


$$P(A \cup B) \neq 0$$

Mutually Exclusive outcomes:



P(rolling 1) =P(rolling 2) = …= P(rolling 6) = 1/6

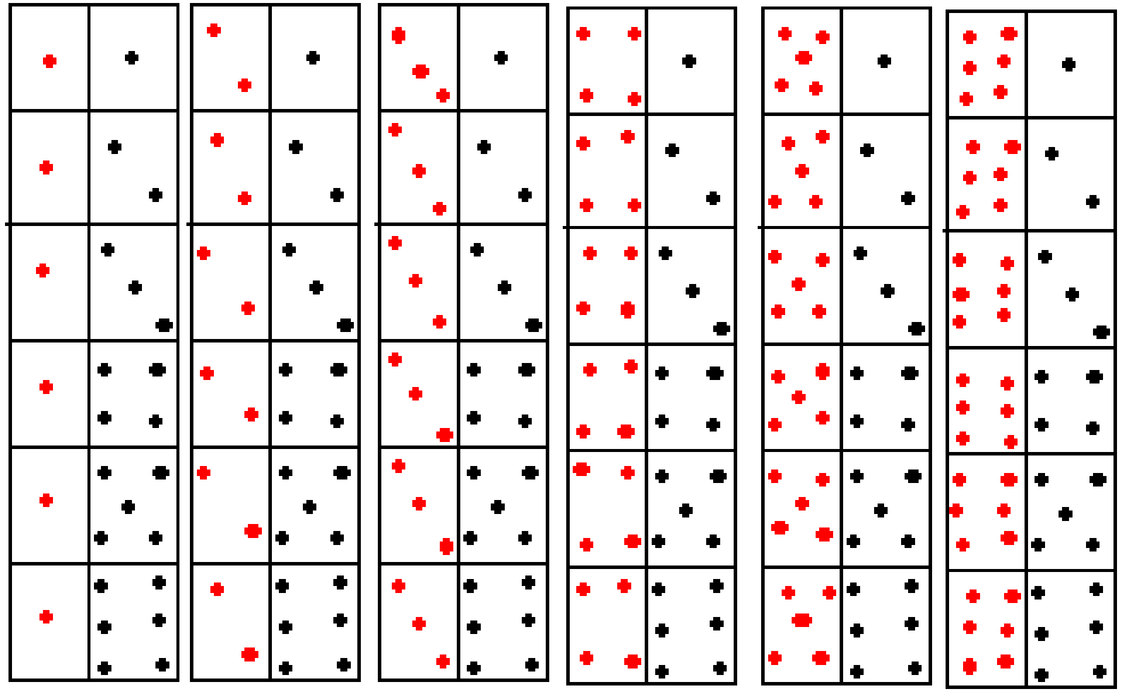


P(rolling 1∩ rolling 1) =P(rolling 1∩ rolling 2) = …= P(rolling 1∩ rolling 6)
= 1/36

ex. Two dice are rolled. The event of interest is the two dice faces add up to 3. What is the probability of this event?

The probability of this event is the sum of the probabilities of the elementary outcomes in this subset.

<u>The full set of elementary outcomes of rolling two dice</u>



P(showing 3 dots)=2 ways=
1/36+1/36 = 2/36

Two dice are rolled and the red die shows 1, what is the probability of rolling a total of 3?
P(total 3|Die 1=1) = 1/6

A <u>fair</u> coin is tossed six times, and the results are recorded in the order that they appear. Which of the following outcomes is most likely to occur.

H=heads; T=tails

1. HHHTTT: _ _ _ _ _ _
2. HTHTHT
3. HTTHTT
4. 1 and 2 are equally likely
5. 1,2 and 3 are all equally likely

For independent events, you can think of them as: _ _ _ _ _ _

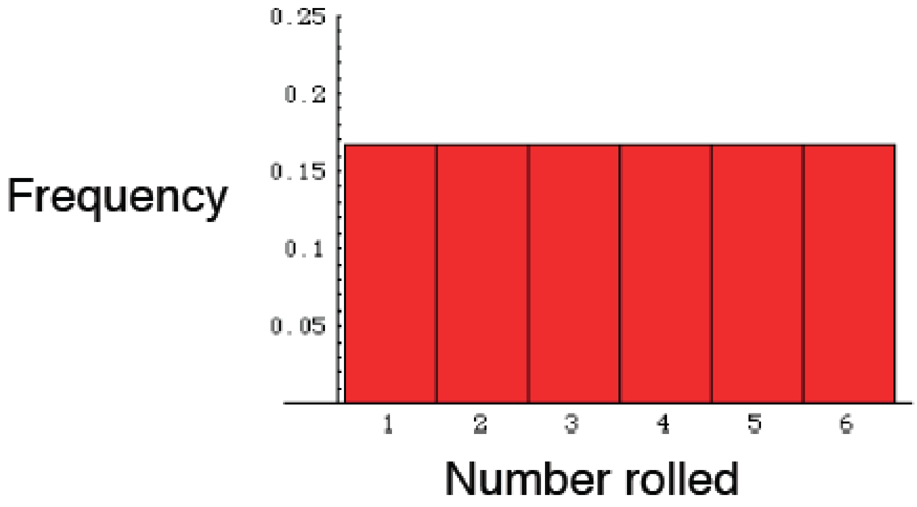
= 0.5*0.5*0.5*0.5*0.5*0.5

Discussion:

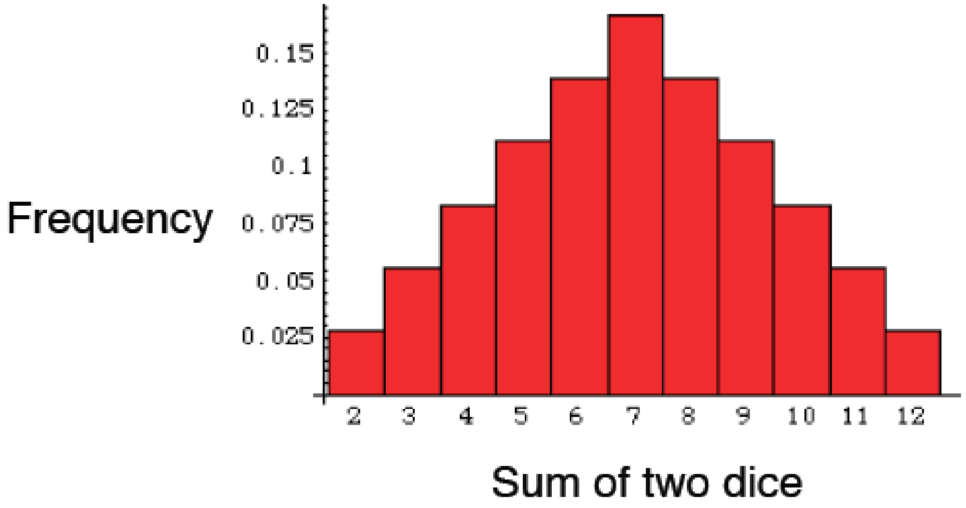
What if a coin is flipped five times and comes up heads each time. Is a tail "due" and therefore more likely than not to occur on the next flip?

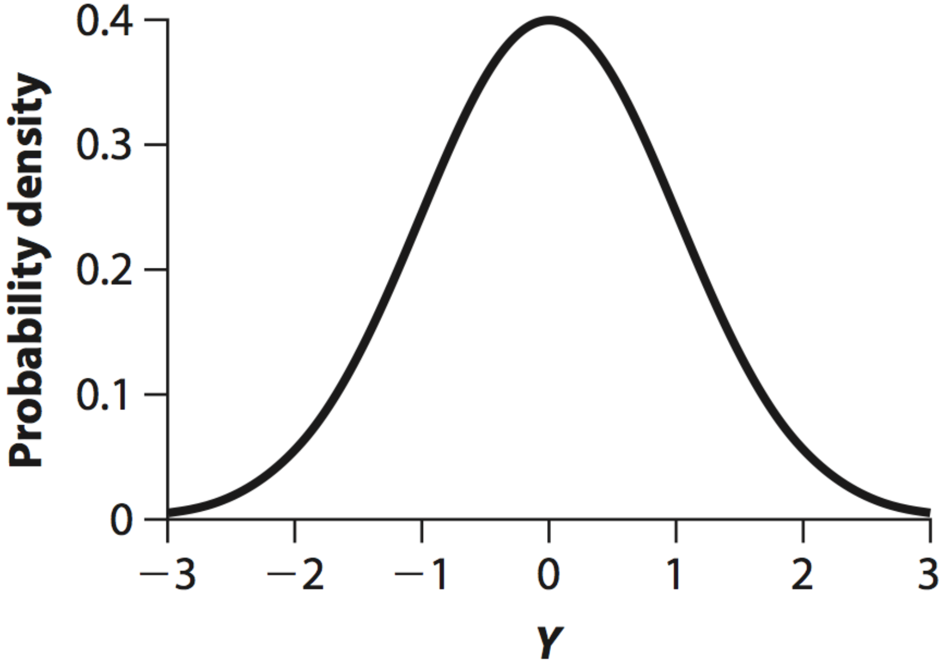
http://onlinestatbook.com/2/probability/gambler.html

Probability distribution for the outcome of a roll of one die:



Probability distribution for the sum of a roll of two dice:





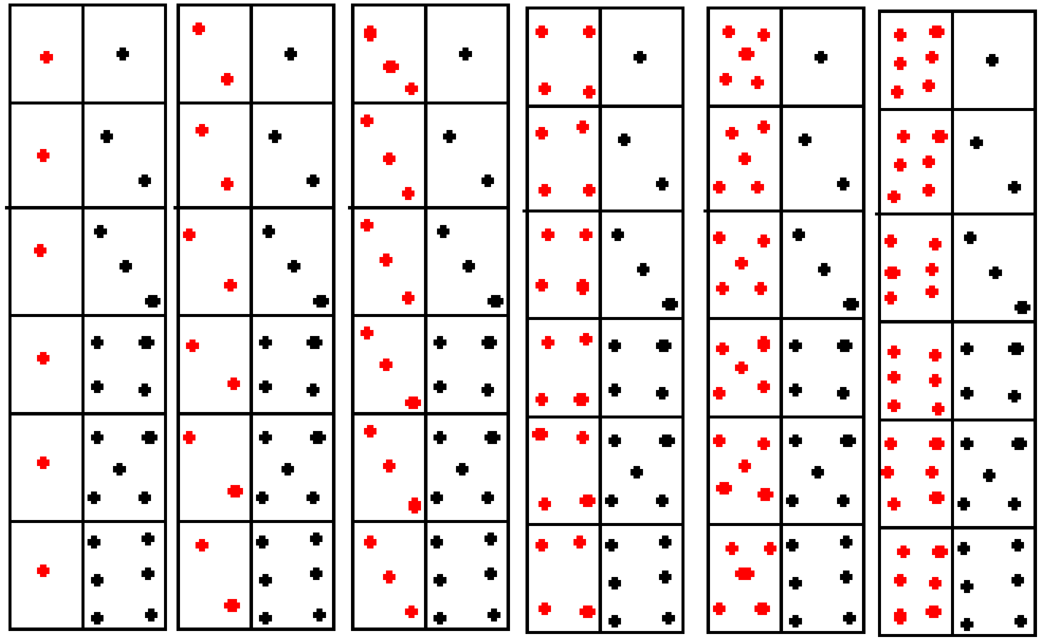
Ex. Event A = Black die is 1, Event B = Red die is 1

*P(A **OR** B) = ?*

a. 6/36
b. 12/36
c. 11/36
d. 1/36

*P(A **AND** B) =?*

a. 6/36
b. 12/36
c. 11/36
d. 1/36



Ex. Event A = Black die is 1, Event B = Red die is 1

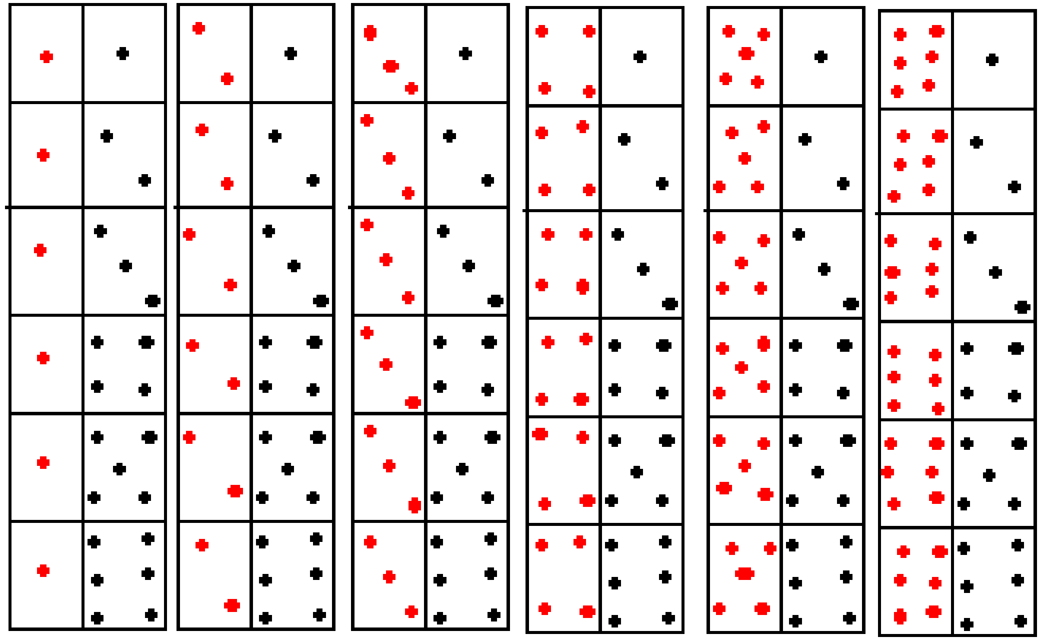
*P(**NOT** A) = ?*

a. 1/36

b. 35/36

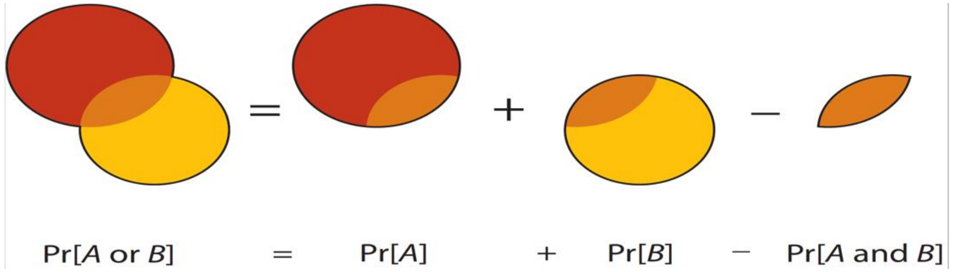
c. 6/36

d. 30/36



Addition Rule:

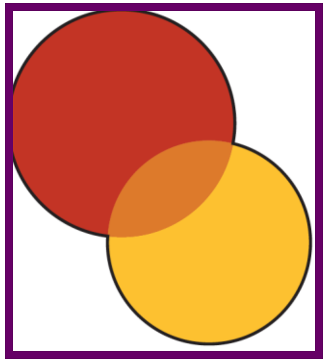
$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

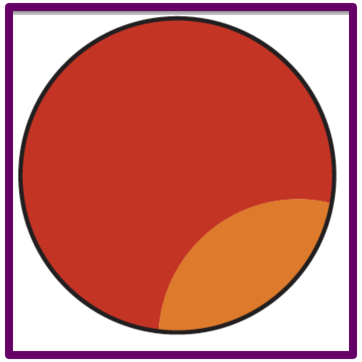


if mutually exclusive: $P[A \cup B] = P[A] + P[B]$

Multiplication Rule:

$$\boxed{P[A \cap B] = P[A \mid B]P[B] = P[B \mid A]P[A] \xrightarrow{IFF independent} P[A]P[B]}$$





- *Two events are independent if the occurrence of one gives no information about whether the second will occur*
- *Two events are dependent if the probability or outcome of one event <u>changes</u> because of the outcome of a second event*

Example: In a forest, imagine that 1% of the trees are infected by a fungal rot and 0.1% have owl nests. What is the probability that a tree has both fungal rot and an owl nest if:

The two events are independent?

a. P(Owls)*P(rot) = 0.01*0.001

b. 0.01+0.001

c. 0.01+0.001-0.01*0.001

The two events are mutually exclusive?

a. 0

b. 0.01*0.001

c. 0.01+0.001

Suppose you take out two cards from a standard pack of cards (52) one after another, without replacing the first card. What is probability that the first card is the ace of spades, and the second card is a heart?

So far, we have learned a great many things about probability:

1. Sample space is made up of elementary outcomes
2. Events can be elementary outcomes or groupings of elementary outcomes
3. Logic operators on probabilities: AND, NOT, OR
4. General Addition rule: $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$
5. IF events A and B are mutually exclusive, then general addition rule collapses into special addition rule: $P(A \bigcup B) = P(A) + P(B)$
6. General Multiplication rule: P[A and B]=P[A|B]xP[B] = P[B|A]XP[A]
7. If events A and B are independent, general multiplication rule collapses into special multiplication rule
   - allows one to test whether or not two events are independent

8. What about if they are not independent?

<u>Example:</u> *Nasonia vitripennis*, a parasitoid wasp, lays eggs in fly pupae; larval wasps then hatch inside, feed on host, and emerge as adults; the males and females then mate on the spot.

***Nasonia* females manipulate sex of their offspring depending on if host fly pupa previously parasitized**.

- If host not yet parasitized, then *Nasonia* lays mainly female eggs and produces only a few males (one male can fertilize multiple females).
- If host already parasitized, then *Nasonia* lays mostly male eggs.

The state of the host encountered by a female and the sex of an egg laid are <u>**dependent variables**</u> (Werren, 1980)

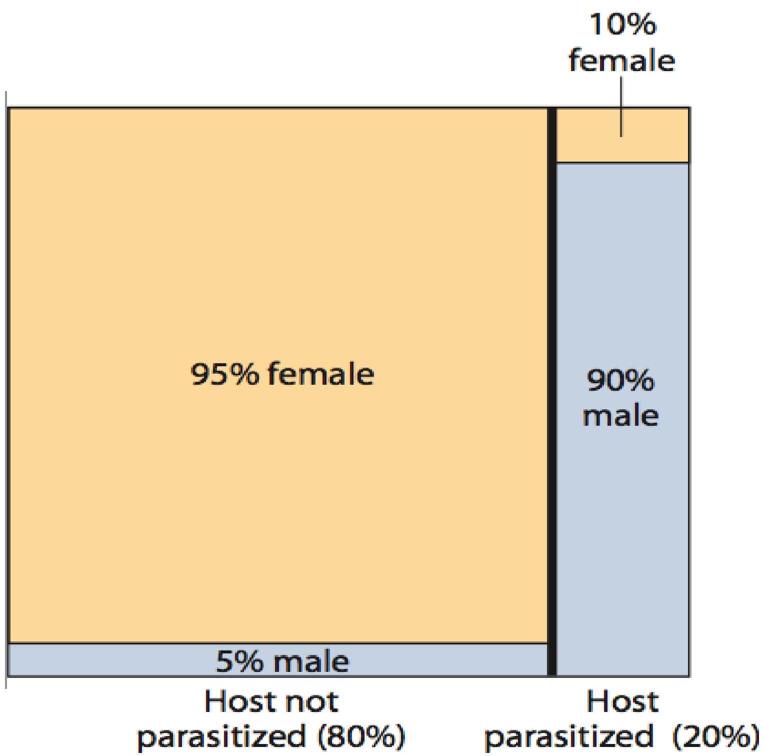
* If host not yet parasitized, then *Nasonia* lays mainly female eggs and produces only a few males (one male can fertilize multiple females).

* If host already parasitized, then *Nasonia* lays mostly male eggs.

State of host (parasitized, not parasitized)

Possibly Dependent variable based on mosaic plot

Sex of egg (male, female)



Example: Offspring of two carriers (Nn x Nn):

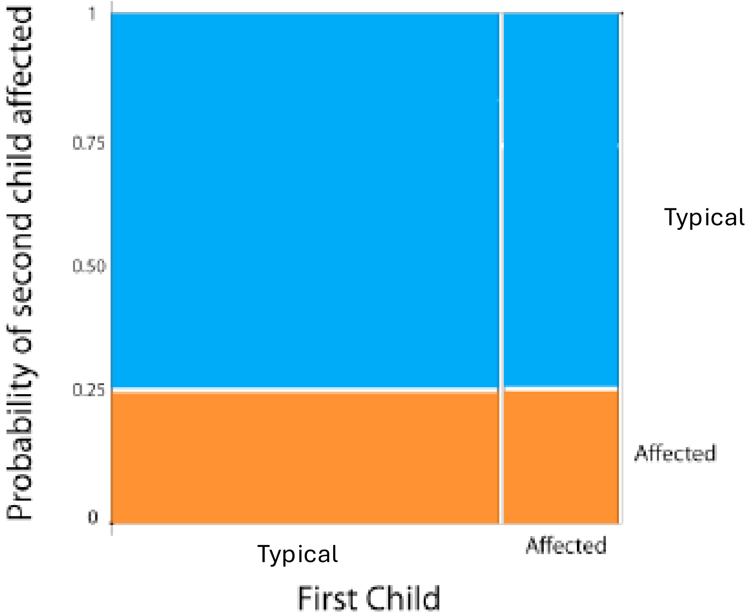
P[night blindness]=0.25

	N	**n**
N	NN	Nn
n	nN	nn

What is the probability that two kids from this family both have night blindness?

P[(1st child night blindness) AND (2nd child night blindness)]
= 0.25 x 0.25 = 0.0625

Possibly Independent variable based on mosaic plot:



Probability trees provide a straightforward method to determine independence or dependence between variables

- Map out probabilities of all mutually exclusive outcomes of variables

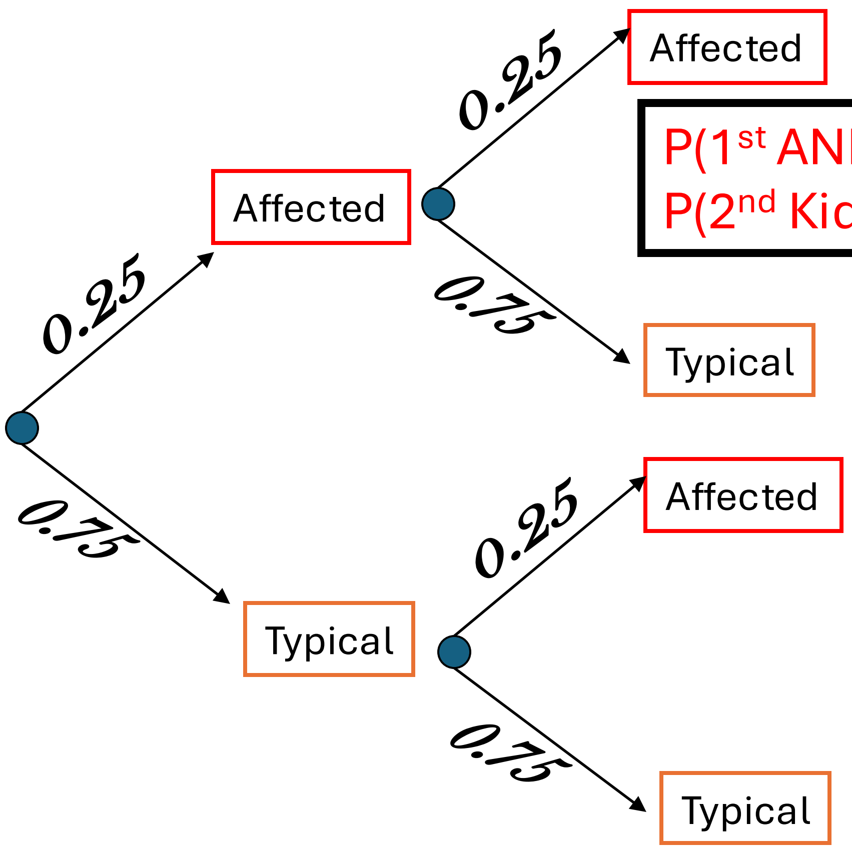
Additional Benefits:

- easy to calculate the probability of any possible outcome sequence for the variables under consideration

- easy to double check that all possibilities have been enumerated

Probability Trees: two events, meiosis is independent

Phenotype of 1^{st} Child | Phenotype of 2^{nd} Child | Probability:



0.25 x 0.25 = 0.0625

P(1^{st} AND 2^{nd} kids affected) = P(2^{nd}|1^{st})P(1^{st})=
P(2^{nd} Kid affected) X P(1^{st} Kid affected)

0.25 x 0.75 = 0.1875

0.75 x 0.25 = 0.1875

0.75 x 0.75 = 0.5625

So far, we have learned a great many things about probability:

1. Sample space is made up of elementary outcomes
2. Events can be elementary outcomes or groupings of elementary outcomes
3. Logic operators on probabilities: AND, NOT, OR
4. General Addition rule: $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$
5. IF events A and B are mutually exclusive, then general addition rule collapses into special addition rule: $P(A \bigcup B) = P(A) + P(B)$
6. General Multiplication rule: P[A and B]=P[A|B]xP[B]
7. If events A and B are independent, general multiplication rule collapses into special multiplication rule
   - allows one to test whether or not two events are independent

8. What about if they are not independent?

A large population of giant pandas has five alleles at one gene labeled: A_1, A_2, A_3, A_4, A_5. They have corresponding frequencies in the population population 0.1, 0.15,0.6, 0.05, 0.1. In this randomly mating population, the two alleles present in any individual are independently sampled from the population as a whole.

a. What is the probability that a single allele chosen at random from this population either **A_1** or **A_4** **= $P(A_1)+P(A_4)$ = 0.1 +0.05 = 0.15**

b. What is the probability that the individual has **two A_1** alleles? $P(A_1)*P(A_1)$ = 0.1*0.1 = 0.01

c. What is the probability that an individual is ***not*** **A_1A_1**? 1-$P(A_1\ A_1)$ = 1-0.01 = 0.99

d. What is the probability, if you drew two individuals at random from this population that neither of them would have an **A_1A_1** genotype?

P(1st diploid isn't A1A1)? 0.99; P(2nd diploid isn't A1A1) = 0.99

0.99*0.99 = 0.9801

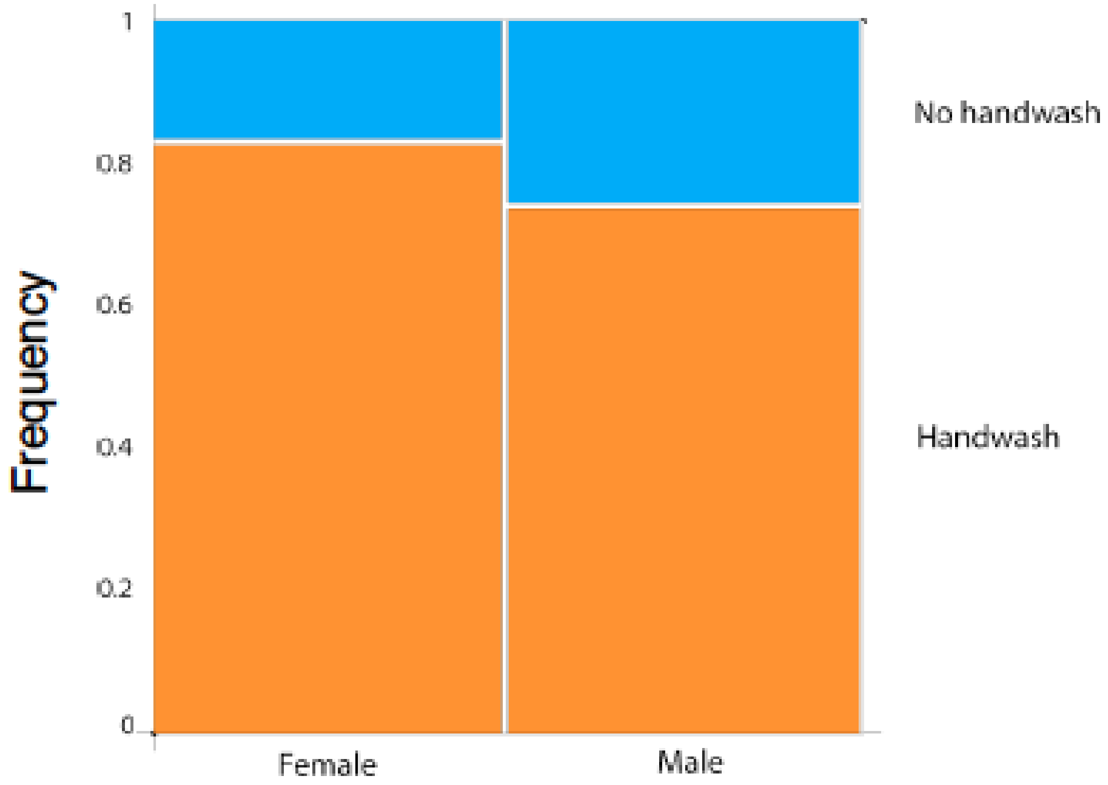
For your certificate of completion:

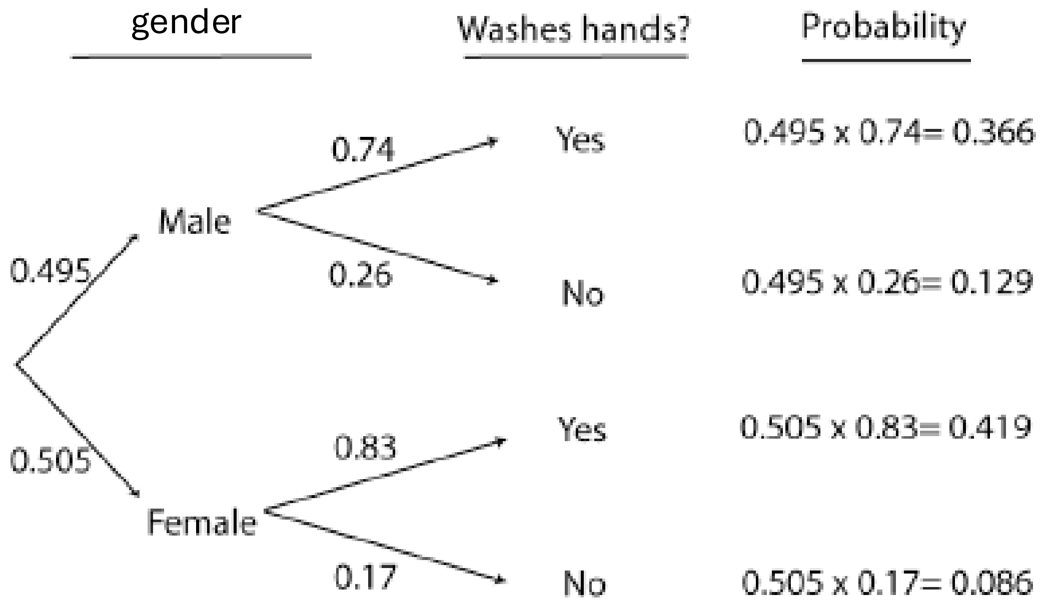
e. What is the probability, if you drew two individuals at random from this population that at least one of them would have an **A_1A_1** genotype?

f. What is the probability that three randomly chosen individuals would have ***no*** **A_2** or **A_3** alleles?

Example: Is washing your hands after using the washroom dependent on gender?

- P[male] = 0.495
- P[male washes his hands]=0.74
- P[female washes her hands]=0.83





Conditional Probability:

The probability that an event occurs given that a condition is met

P[X|Y] = P[X and Y]/P[Y]

This is read as “the probability of X given Y”

It means: the probability of X if Y is true

Fancier way of writing the total probability of an event:

$$\boxed{P[X] = \sum_{Y} P[X \mid Y] P[Y]}$$

Conditional Probability: **P[X|Y] = P[X and Y]/P[Y]**

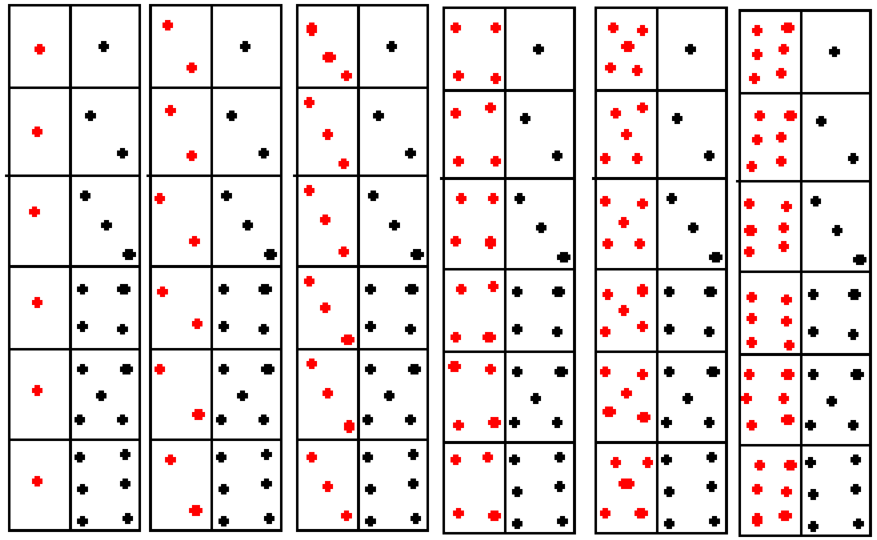
Example:

What is the probability that two dice will sum to three?

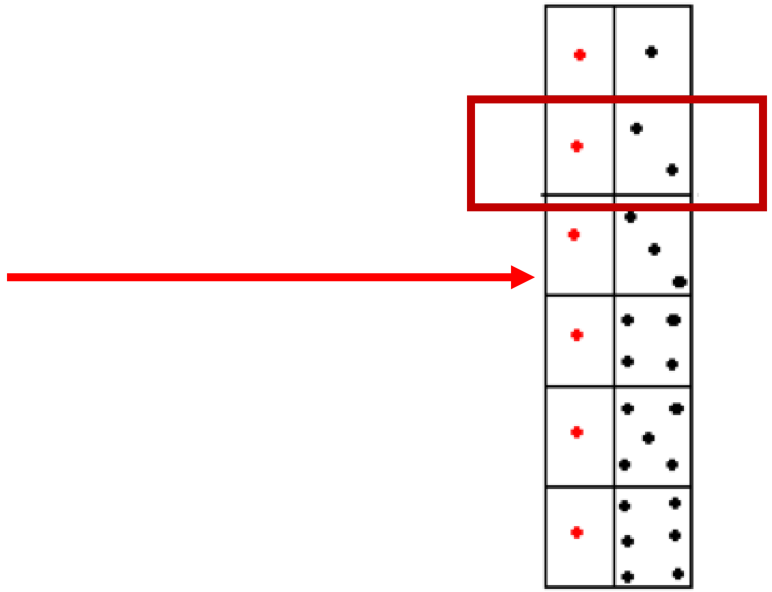
*-this is really asking **P[X and Y]** where $\boldsymbol{X}_{red\ die}$ =1 or 2 and $\boldsymbol{Y}_{black\ die}$ = 1 or 2*

Now: **what if we already have rolled the first die and know that we have a one? Event $\mathbf{X_{red}}$=1**

Reduced state space, from 36 to 6:



P[Sum to three] =2/36



P[Sum to three]=1/6

- **<u>Very</u>** important to understand conditional probability before we tackle Bayes'

- Conditional probability can be a little confusing; sometimes using a Venn diagram with **3** events instead of 2 makes it clearer (example 5 and 6):

  https://www.nagwa.com/en/explainers/403141497934/

- Some learners have struggled with the difference between $P(A \cap B)$ and P(A|B)

- **Remember:** $\boxed{P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)}$

1st Variable 2nd Variable $P(A \cap B)$ **Dependent or Independent?**