

Probability

Last updated: Wednesday 15th July, 2015 17:23

Let us get the intuitive idea behind the probability before we get to the formal definition.

Monty Hall Problem

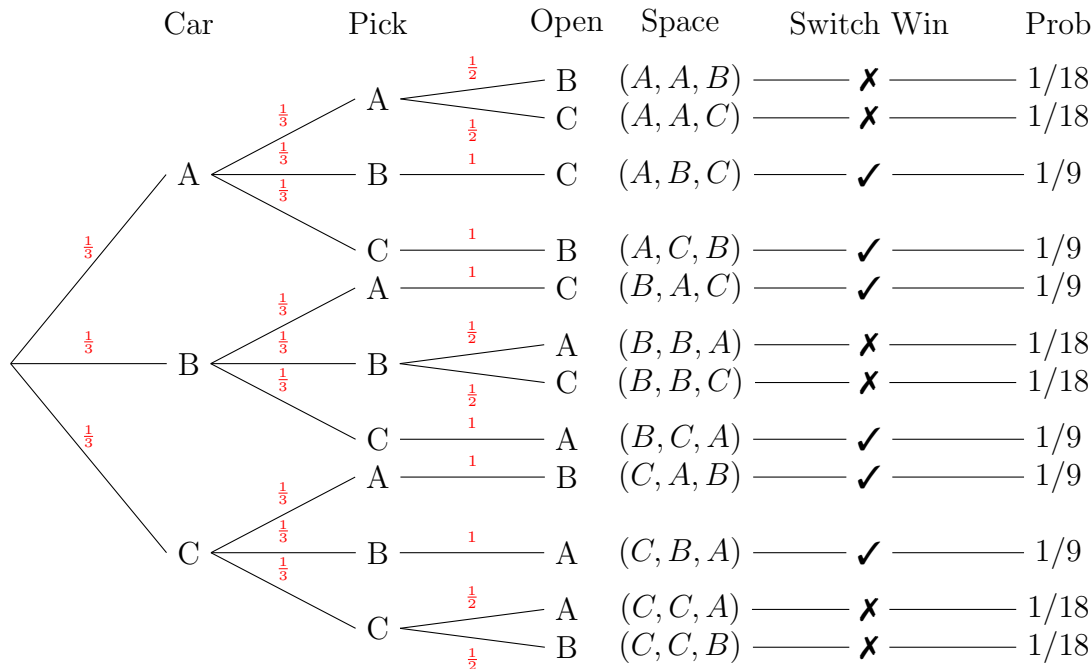
The first example is based on an American game show in 1980's called . It is called Monty Hall problem named after the host of the game show Monty Hall.

The game show goes like the following: you are given three doors to choose. There is a car behind one of the door and behind the other two door has goats behind it.

- First, the player pick a door.
- Since there are two empty doors, at least one empty door is not picked. The host will then open one of the empty door that is not picked by the player.
- The host then ask the player whether to switch the door he pick to the other non open door or not.

The question is should the player switched or not.

This problem can be explained easily by just drawing a tree.



From the picture, to calculate the probability that you switch and win, all we need to do is add up all the probability of the sample that you win.

$$\begin{aligned}
 \Pr(\text{switch and win}) &= \Pr([A, B, C]) + \Pr([A, C, B]) + \Pr([B, A, C]) + \\
 &\quad \Pr([B, C, A]) + \Pr([C, A, B]) + \Pr([C, B, A]) \\
 &= 6 \times \frac{1}{9} = \frac{2}{3}
 \end{aligned}$$

Thus, the probability of switch and win is $\frac{2}{3}$ and the probability of not switch and win is $\frac{1}{3}$. So, you should switch.

One way to understand this seemingly counterintuitive result is that if we switch the door we are essentially picking two doors. So, the chance should really be twice as much.

The confusion usually comes from the fact of what is being define of equal probability. Naively one would think that each outcome have the same probability associated with it. But, in fact, each outcome do not have the associated probability as you can see from the tree we drew.

Strange Dice

Let us consider another example. The three dices have the following faces

Dice A = 2, 6, 7

Dice B = 1, 5, 9

Dice C = 3, 4, 8

The rule of the game is simple. There are two players each one pick a dice who ever roll greater number wins.

By drawing a tree(I would if i have time), one can conclude

- A wins B with probability of 5/9
- B wins C with probability of 5/9
- ans surprisingly C wins A with probability of 5/9

This show that probability does not have the associative probability.

Formal Definition

Def: A countable(just means the set where you can list the members) sample space S is a non-empty countable set. An elements $\omega \in S$ is called an outcome.

For example, for tossing a coint two times, the same space would be

$$S = \{HH, HT, TH, TT\}$$

Def: A probability function is a function $Pr : S \rightarrow \mathbb{R}$ such that

- $Pr[\omega] \geq 0 \forall \omega \in S$. Probability associated with an outcome cannot be negative.
- $\sum_{\omega \in S} Pr[\omega] = 1$. The sum of all probability is 1.

The probability for the two coin toss, the probability function would be

$$Pr[HH] = Pr[HT] = Pr[TH] = Pr[TT] = 0.25$$

Def: A probability space is the sample space together with probability function.

For example, the probability space would be (S, Pr) where

$$S = \{HH, HT, TH, TT\}$$

and the probability function

$$Pr[HH] = Pr[HT] = Pr[TH] = Pr[TT] = 0.25$$

Def: An event is a subset of a samplespace.

$$E \subseteq S$$

For example, an event where the two coin toss give the same face

$$E = \{HH, TT\}$$

Def: We can then define the probability on the event as the sum of all the outcome in the event.

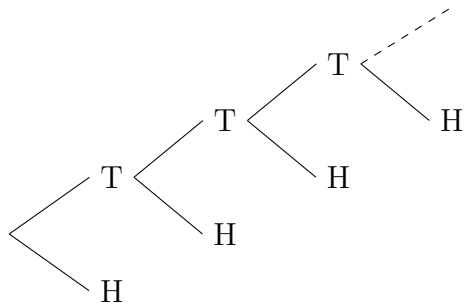
$$Pr[E] = \sum_{\omega \in E} Pr[\omega]$$

For example, the probability that the two tosses come up the with the same face is

$$Pr[\{HH, TT\}] = Pr[HH] + Pr[TT] = 0.5$$

These definitions match with what our intuition says about probability. We can retrace what we did in the montehall case and the strange dice case to these definitions.

The sample space may not be a finite set. For example, let us consider a game of two player fair coin toss. The two players alternate the coin toss. Whoever get the first head wins. For instance, if the first player get head the first time he toss. He just win and the game ends.



The sample space for this game would be

$$S = \{H, TH, TTH, TTTH, TTTTH, \dots\}$$

The sample space is infinite. The probability of each one looks like $\Pr[H] = 1/2$, $\Pr[TH] = 1/4$, $\Pr[TTH] = 1/8$, etc.

Now we want to calculate the probability that the first player wins. The event that the first player win is

$$E = \{H, TTH, TTTTH, TTTTTH, \dots\}$$

This means the probability of the first player winning is

$$\begin{aligned} \Pr[E] &= \Pr[H] + \Pr[TTH] + \Pr[TTTTH] + \dots \\ &= \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \frac{1}{2^7} + \frac{1}{2^9} + \dots \\ &= \frac{1}{2} \left(1 + \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \frac{1}{2^8} + \dots \right) \\ &= \frac{1}{2} \left(1 + \frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \frac{1}{4^4} + \dots \right) \\ &= \frac{1}{2} \left(\frac{1}{1 - \frac{1}{4}} \right) \\ &= \frac{2}{3} \end{aligned}$$

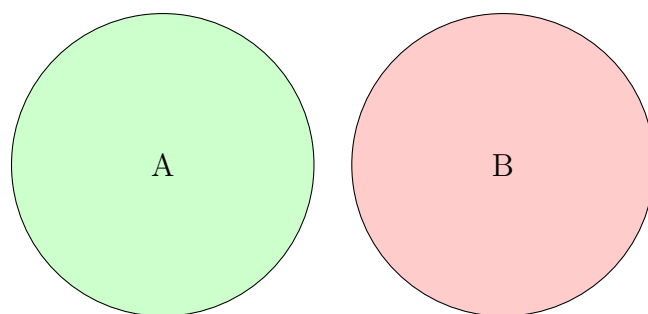
Common sense rule.

Since the probability of an event is just a linear sum of probability of all outcome in the event set. A lot of cardinality rule we learn from counting works here too.

Sum Rule

For example, if E_1 and E_2 are disjoint ($E_1 \cap E_2 = \emptyset$) then

$$\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2]$$



This is quite intuitive. Suppose I give you 1 yoyo at random. You have certain probability of getting a red yoyo and a certain probability of getting pink yoyo. The probability of getting pink yoyo or red yoyo is then just the sum of the two probability.

Complement Rule

The probability of something not happens is just one minus the probability of that event happens.

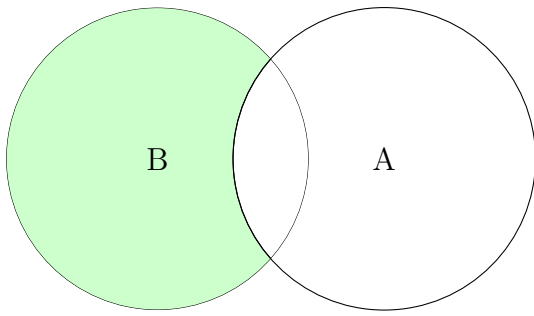
$$\Pr(\bar{E}) = 1 - \Pr(E)$$

For example, the probability of not raining today is just one minus the probability of raining.

Difference Rule

The probability of B happens but not A is just the probability that B happens minus the probability that both A and B happens.

$$\Pr(B - A) = \Pr(B) - \Pr(B \cap A)$$



Union Rule

All the inclusion-exclusion rule works here too. For example,

$$\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2] - \Pr[E_1 \cap E_2]$$

Conditional Probability

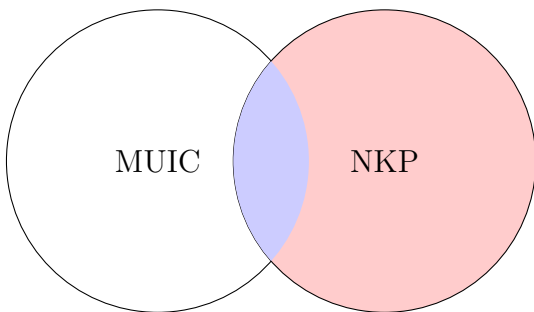
Most of the time the probability that you are interested in is the probability of something happens given some information. Consider picking a random person in Thailand. The probability of that person being an MUIC student would be

$$\Pr(\text{MUIC Student}) = \frac{\# \text{ MUIC Student}}{\# \text{ People in Thailand}}$$

But, now if we ask that person where does he/she lives and the answer is Nakornpathom, then it is clear the the probability that he/she is an MUIC student would be much higher. So, this kind of probability is called conditional probability: the probability of something given some information. Specifically, the probability of some one being an MUIC student given that he/she lives in Nakornpathom(NKP) is denoted by

$$\Pr(\text{MUIC Student} | \text{Live in Nakornpathom})$$

We can visualize that probability as



This means that the probability we want to calculate is the number of people who lives in the NKP and study at MUIC (blue region) normalized by all the people live in Nakornpathom (blue and red region).

$$\Pr(\text{MUIC} | \text{NKP}) = \frac{\# (\text{MUIC} \cap \text{NKP})}{\# \text{ NKP}}$$

With this notion we can define the conditional probability as

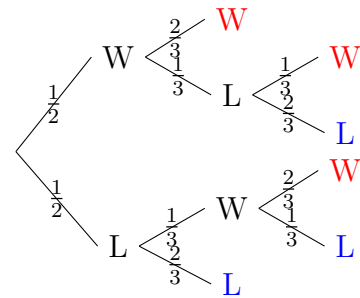
Def: The probability of event A given event B , $\Pr(A|B)$, is given by

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

This is quite important law in probability that it is given a name Bayes's rule.

Soccer Match

Let us look at an example. Suppose you are in a soccer game, you have to win 2 out of 3 games to win. The probability of you winning the first game is $1/2$. But, if you win one game the next game the probability of winning again is $2/3$ due to mysterious morale boost. From this situation you can draw the tree as



From the picture above, the probability of winning is

$$\Pr(\text{Winning}) = \frac{1}{3} + \frac{1}{18} + \frac{1}{9} = \frac{1}{2}$$

But we want to know more. If we happen to win in the first match(WFM), we want to know the probability that we will win the tournament(WT). The probability that we want to calculate is, by Bayes' rule,

$$\begin{aligned}\Pr(WT|WFM) &= \frac{\Pr(WT \cap WFM)}{\Pr(WFM)} \\ &= \frac{\frac{1}{3} + \frac{1}{18}}{\frac{1}{2}} \\ &= \frac{7}{9}\end{aligned}$$

You can do a similar thing and find that the probability of losing the tournament given that you win the first round is $\frac{2}{9}$.

We can ask *a posteriori* question, which is a fancy word meaning reverse in time. For example, we can ask the question that given that we in the tournament, what is the probability that we win the first round. We can also calculate this using Bayes's rule:

$$\begin{aligned}P(WFM|WT) &= \frac{\Pr(WT \cap WFM)}{\Pr(WT)} \\ &= \frac{\frac{1}{3} + \frac{1}{18}}{\frac{1}{3} + \frac{1}{18} + \frac{4}{18}} \\ &= \frac{7}{11}\end{aligned}$$

Tree Diagram

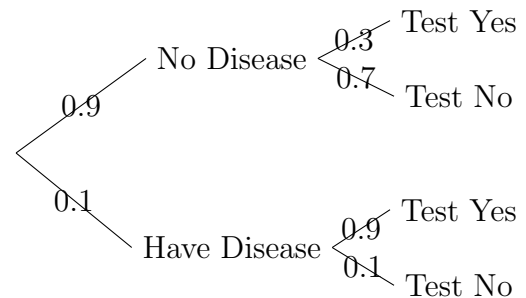
The bayes rule is also the reason by the tree diagram we drew works. We try to compute the probability that the all the event in the branch leading to the leaf is. The probability that we write on each edge is just the probability that something happens given all the other things leading to that edge happen. For example, we do this without even thinking about it

$$\begin{aligned}\Pr(A \cap B) &= \Pr(B) \cdot \Pr(A|B) \\ \Pr(A \cap B \cap C) &= \Pr(C) \cdot \Pr(B|C) \cdot \Pr(A|B \cap C)\end{aligned}$$

Medical Trial

Let us consider a medical test for a disease that have the following sensitivity and specificity.

- **Sensitivity.** If you have the disease, then there is 90% chance that the test will tell you that you have the disease. This is also called a true positive rate.
- **Specificity.** If you *do not* have the disease then there is a 70% chance that the test will tell you that you *do not* have a disease. This is also called the true negative rate.
- 10% of the population have the disease.



An important question to ask is that if you get the test and the test says you are positive(PT) what is the probability that you do have the disease(D).

$$\begin{aligned}\Pr(D|PT) &= \frac{\Pr(D \cap PT)}{\Pr(PT)} \\ &= \frac{0.1 \times 0.9}{0.1 \times 0.9 + 0.9 \times 0.3} = 25\%\end{aligned}$$

So, even though you should be a bit worried more than when you didn't have the test(10%). You should not be too much worried since the probability of you having the disease given that the test is positive is just 25%.

Let us calculate the probability of the test being correct.

I can be correct more often just just keep saying no. But that is not a very useful test. So, the "correctness" is not the measure of how useful a test is. There is really no right answer for this. It is all depend on what is the cost of the false negative and the cost of false positive. If the disease you are diagnosing is an contagious zombie disease then false negative would cost you a lot: end of humanity while false postive doesn't cost much. Whereas if the disease we are talking

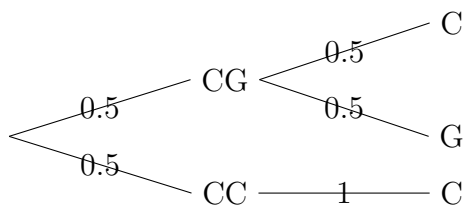
about is just a mild disease that we can fix when the true symptom occur but it costs a lot money to cure, then one false positive would cost a lot, while false negative doesn't really matter we can just wait until the symptom occur and fix it.

Yoyo Box

Suppose I have 2 boxes of yoyo.

- One box has a cola yoyo and a grape yoyo. The other box has 2 cola yoyo.
- I then pick a box at random.
- Then, we randomly draw a yoyo from the box we chose.
- If the yoyo we drew is a grape yoyo, then we randomly pick a box again.
- If the yoyo we drew is a cola, then we play the game by drawing the other yoyo from the box. If it comes out grape you win, and if it comes out cola I win.

We can analyze the game again using the tree diagram. What we want to find is the probability that the box we pick is the one with two flavors given that the first yoyo we drew is a cola.



Of course, none of the game you play in this class is fair. AJ always cheat. The probability of the box we pick is CG(Cola Grape) given that the yoyo we draw is a Cola is

$$\begin{aligned}\Pr(CG|C) &= \frac{\Pr(CG \cap C)}{C} \\ &= \frac{\frac{1}{4}}{\frac{3}{4}} \\ &= \frac{1}{3}\end{aligned}$$

This means AJ wins with the probability of $\frac{2}{3}$. This is not a fair game. As an exercise you should think about whether having 2 CG boxes and 1 CC box make the game a fair game.

Conditional Probability Properties

These can be understand quite easily by just drawing a diagram.

$$\Pr(A) = \Pr(A|E) \cdot \Pr(E) + \Pr(A|\bar{E}) \cdot \Pr(\bar{E})$$

This reads the probability of event A is the sum of the probability of event A happen given E times probability that E happens and the one where E doesn't happen. This is called the law of total probability.

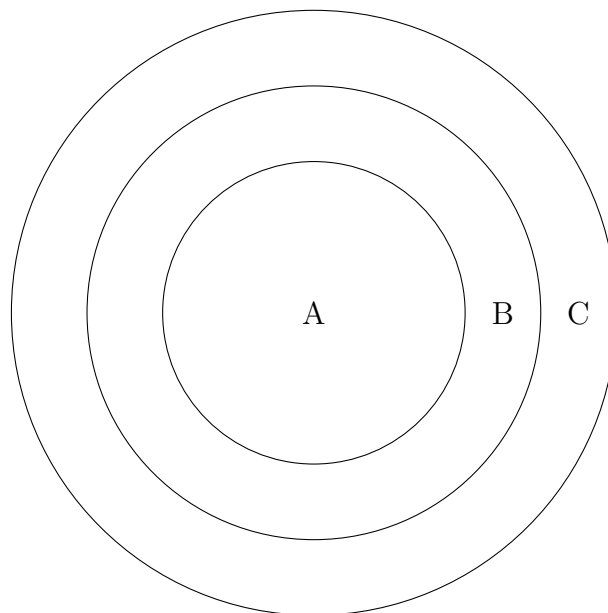
Also, you can conditioning on the intersection and it works like it should since we just change the normalization.

$$\Pr[A \cup B|C] = \Pr[A|C] + \Pr[B|C] - \Pr[A \cap B|C]$$

However, one need to be careful about conditioning on the union.

$$\Pr[C|A \cup B] \neq \Pr[C|A] + \Pr[C|B]$$

A counterexample can be shown using the figure below



From the picture $\Pr[C|A] = 1$ since C is guarantee to happen if A happen. Also, $\Pr[C|B] = 1$ for the same reason. So, if we add the two up we will get 2 which is not a probability.

Simpson Paradox

Consider the following table for the ontime rate for American Airlines(AA) and American West Airlines(AW). The question is which one should you fly?

	AA		AW	
LA	500/560	89%	700/800	87%
PHO	220/230	95%	4900/5300	92%
SD	210/230	92%	400/450	89%
SF	500/600	83%	320/450	71%
Seattle	1900/2200	86%	200/260	77%
Total	3300/3820	87%	6520/7260	90%

The peculiar thing about this table is that even though AA is better than AW in every single airport the total rate for AA is worse than AW. You can ask the question that given that you will be flying off from LA, what is the airline we should choose? Clearly it is AA. You can ask this question for every single airport, the answer will always be AA.

However, if you do not know where you are flying off, you will be better off with AW.

This results is called Simpson's Paradox. Newspaper do this all the time. But, since you have learned this, I hope whenever you see someone totalling the ratio, you will notice the pitfall.

Independence

Def: Event A and event B are independent iff

$$\Pr(A|B) = \Pr(A) \text{ or } \Pr(B) = 0$$

In other word, given the information about B does not tell you anything about A .

For example, let us consider tossing two fair coins consecutively. Let us consider two events

- Event A . The two coins are the same.
- Event B . The first coin is head.

Are A and B independent?

All we need to do is checking whether $\Pr(A|B)$ and $\Pr(A)$ are equal. The probability that the coins are the same($\Pr(A)$) is $1/2$. If the

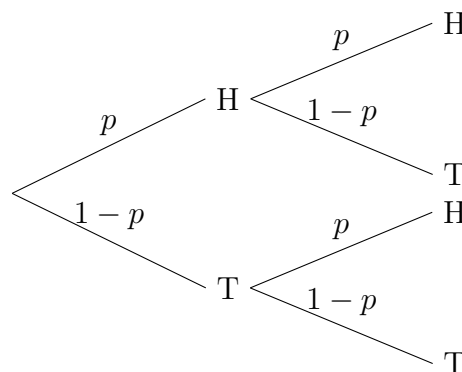
first coin is head then the probability that the two coins are the same is the probability that the second coin is also head so $\Pr(A|B) = 1/2$. So, event A and event B are independent.

It should also be noted that using fair coin doesn't imply the fact that the result of the two coins are independent. You can tape the coins together. The two events are still independent.

What if the coin is not fair? This means that the probability of getting head is p and the probability of getting tail is $1 - p$. Would the two event still be independent.

- Event A . The two coins are the same.
- Event B . The first coin is head.

We can draw the tree:



All we need to compare is $\Pr(A|B)$ and $\Pr(A)$.

- $\Pr(A|B) = p$
- $\Pr(A) = p^2 + (1 - p)^2$

So, the two events are not independent. The reason is quite simple. Suppose that the coin land on heads way more often than tail. If the first coin lands on head, since the second coin lands on head way more often than the tail that means the chance of you getting a match is higher. This means that the information about the first coin tells give you information about the second event.

Theorem: (Product Rule) If event A is *independent* of event B then

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B)$$

Some book use this as the definition of independence. The two definiton are equivalent.

Proof: There are two cases

1) If $\Pr(B) = 0$ then $\Pr(A \cap B) = 0 = \Pr(A) \Pr(B)$ since B never happen.

2) If not then we know that

$$\Pr(A \cap B) = \Pr(B) \Pr(A|B) \underset{\substack{\uparrow \\ \text{indep}}}{=} \Pr(B) \Pr(A)$$

□

Theorem: Independent is symmetric. If event A is independent of event B then event B is independent of event A .

Proof: Left for the reader as an exercise □

Mutual Independence

Let us consider consider tossing 3 fair independent coin. Let

- A be the event that the 1st and the 2nd coin match.
- B be the event that the 2nd and the 3rd coin match.
- C be the event that the 1st and the 3rd coin match.

Let us see if A and B independent. This is quite easy to see since $\Pr(A) = 1/2$ and $\Pr(B) = 1/2$. $\Pr(A \cap B) = \Pr(HHH) + \Pr(TTT) = 1/4$. So,

$$\Pr(A \cap B) = \Pr(A) \Pr(B)$$

So that means A and B are independent. We can do the same thing for other pairs.

- A and B are independent $\Pr(A \cap B) = \Pr(A) \Pr(B)$
- B and C are independent $\Pr(B \cap C) = \Pr(B) \Pr(C)$
- A and C are independent $\Pr(A \cap C) = \Pr(A) \Pr(C)$

The question is can we then conclude that the probability that all three event happens is the product of each probability?

$$\Pr(A \cap B \cap C) = \Pr(A) \Pr(B) \Pr(C)$$

This is not the case since the event that all three event happens is the event that all three coins are the same that means the $\Pr(A \cap B \cap C) = 1/4$. But the product of the three is $1/8$. So,

$$\Pr(A \cap B \cap C) = \frac{1}{2} \neq \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}$$

So, we learn a very important thing that pair wise independent does not imply that the three are *mutually independent*. This leads us to define a stronger assumption.

Def: Event A_1, A_2, \dots, A_n are *mutually independent* iff $\forall i \forall J \subseteq \{1, 2, \dots, n\} - \{i\}$,

$$\Pr(A_i | \bigcap_{j \in J} A_j) = \Pr(A_i).$$

The symbol looks a bit scary but all it reads is that the probability that any event A_i happens is independent to all possible intersection of all the other events. This definition is a bit hard to work with so let us use another definition.

Def: (Equivalent to above) Event A_1, A_2, \dots, A_n are *mutually independent* iff $\forall J \subseteq \{1, 2, \dots, n\}$

$$\Pr\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \Pr(A_j)$$

This just say that the probability of any intersection of events is just the product of the probability.

This means that if you need to check for *mutually independent* you need to check all the intersection. Specifically, this means for three events A_1, A_2, A_3 we need to check for all of the following

$$\Pr(A_1 \cap A_2) = \Pr(A_1) \Pr(A_2)$$

$$\Pr(A_1 \cap A_3) = \Pr(A_1) \Pr(A_3)$$

$$\Pr(A_2 \cap A_3) = \Pr(A_2) \Pr(A_3)$$

$$\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1) \Pr(A_2) \Pr(A_3)$$

You cannot skip any of these.

Let us look at another example of tossing two independent dice.

405 • Let A be the event that the first dice is 1,2, 412 $\Pr(A \cap B \cap C) = \Pr([3, 6]) = 1/36$. $\Pr(A) = 1/2$
 406 or 3. 413 $\Pr(B) = 1/2$ $\Pr(C) = 4/36 = 1/9$. That means
 407 • Let B be the event that the first dice is 3,4, $\Pr(A \cap B \cap C) = \frac{1}{36} = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{9}$
 408 or 5.
 409 • Let C be the event that the sum of the two 414 But, does this means that A , B and C are
 410 dice is 9. 415 pairwise independent. Unfortunately, no. $\Pr(A \cap$
 411 Let us first check the last condition whether 416 $B) = 1/6$, $\Pr(A) = 1/2$, $\Pr(B) = 1/2$
 $\Pr(A \cap B) \neq \Pr(A) \Pr(B)$
 $\Pr(A \cap B \cap C) = \Pr(A) \Pr(B) \Pr(C)$. 417 So, A , B and C are not mutually independent.