General Probability

Que 1) Given two fair dices, what is the probability that two dices sum to 8? What is the probability that two dices sum to 8 when the first dice is 3?

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

There are 36 (6*6) outcomes for tossing two fair dices, and the outcomes when two dices sum to 8 are:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

(2, 6), (3,5), (4,4), (5,3), (6,2);

The probability of two dices sum to 8 is 5/36.

For the second part, it is a conditional probability that we are calculating. Assume event A is two dices sum to 8, and event B is the first dice is 3. We know that event B's outcomes are:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

(3,1), (3,2), (3,3), (3,4), (3,5), (3,6)

and only (3,5) makes event A happen, thus the probability is 1/6.

Que 2) Person A and Person B are playing archery together. Assume their abilities to fire the arrow at the target are exactly the same, and the probability of getting the target is 0.5 for both of them. Now that given A has fired 201 arrows and B has fired 200 arrows, what is the probability that A gets more targets than B?

201 is not an even number so that let's consider 200 games first. Assume event A is A fires more arrows on target than B in 200 games, and event B is B fires more arrows on target, and C is that they fire equal amount of arrows on targets. We have:

$$P(A) + P(B) + P(C) = 1$$

Since A and B perform equally at archery, for 200 games, we have P(A) = P(B). Thus:

$$2P(A) + P(C) = 1$$

Now move to the extra game that person A plays. If in the last 200 games:

- A is higher than B, then no matter A fires on target or not for this extra game, A is still higher than B.
- If A is lower than B, even if A fires on target for the extra game, we would observe the most A=B, and A will still not be over B.
- If A=B, if A fires on target for the extra game, then A will be higher than B, and the probability that A shoot on target for any game is 0.5.

Thus, the total probability that A is higher than B is:
$$P(A) + 0.5 * P(C)$$

We know that 2P(A) + P(C) = 1, if we divide 2 on both sides, we will have:
$$P(A) + 0.5 * P(C) = 0.5$$

The probability that A gets more targets than B when A plays 201 games and B plays 200 games is 0.5.

Que 3) During flu season, for a two-parent heterosexual family, suppose the probability that at least one parent has the flu is 17%; the probability that the father has the flu is 12%; the probability that both the parents have the flu is 6%, what is the probability that the mother has flu?

Suppose the probability that the father has flu is P(F), and for the mother is P(M). We know:
$$P(M \cup F) = 17\%$$
$$P(F) = 12\%$$
$$P(M \cap F) = 6\%$$
according to the general addition rule of probability:
$$P(M \cup F) = P(M) + P(F) - P(M \cap F)$$
Thus P(M)=11%.

Que 4) You have 40 cards in four colors, 10 reds, 10 greens, 10 blues, and ten yellows. Each color has a number from 1 to 10. When you pick two cards without replacement, what is the probability that the two cards are not in the same color and not in the same number?

We can first calculate the probability of getting two cards with the same number and the probability of getting two cards with the same color, then use one minus the sum of the two probabilities.

The probability of getting two cards with the same number is:

$$P(SameNumber) = \frac{40}{40} * \frac{3}{39} = \frac{3}{39}$$

You can get any number at the first draw, and it doesn't matter. Thus the first draw doesn't affect the probability, but for the second draw, you only have 39 cards left, and you need to pick the same number as the first draw. For each number, there are four cards with the same number in different colors. Thus for the second draw, you can only pick 3 out of 39 cards.

Same logic for getting two cards with the same color:

$$P(SameColor) = \frac{40}{40} * \frac{9}{39} = \frac{9}{39}$$

We pick any color in the first draw and can only choose 9 cards in the same color out of the 39 left-over cards. The probability of not getting the same number AND same card is:

P = 1-P(Same Number)-P(Same Color) = 27/39

Binomial Distribution

Que 5) Team A and B are playing a game that they have to win 4 out of 7 rounds to win the game. The probability of A wins is p, then the probability of B wins is 1-p (no chance of a tie), what is the probability that they will play all seven rounds? What if the probability of A wins is different in the home field (p) and in the visiting field (q)?

If two teams play all 7 rounds, then for the first 6 rounds, both A and B have to win exactly 3 times, and we don't care who wins at the last round. We can

consider each round as a Bernoulli trial, then the number of times A wins in the first 6 games follows a binomial distribution Bi(n,k,p) with n=6, k=3, and p=p, the probability of A wins. According to the Binomial distribution, the probability of A wins 3 times out of 6 games is:

$$Bi(6,3,p) = \binom{6}{3} * p^3 * (1-p)^3 = \frac{6!}{3! * (6-3)!} * p^3 * (1-p)^3 = 20p^3(1-p)^3$$

Note that team A wins 3 times automatically set team B to win 3 times.

If the two teams have different winning rates playing at home and playing away, we can assume Team A's probability of winning at home is p, away is q, and Team A had won x games at home. The probability of both teams play all 7 rounds will be a function of x, p, and q. Specifically, we know both team A and B have to win 3 rounds, then Team A has to win x games at home, and win 3-x games away; Team B has to win 3-x games away (Team A's home is Team B's visit site, Team B wins aways when A loses at home), and win x rounds at home. The probability of playing seven rounds is:

$$\binom{3}{x} * p^x * (1-p)^{3-x} * \binom{3}{x} * q^{3-x} * (1-q)^x$$

If we have more information about the distribution of x, we can get more information about the probability.

Que 6), Eight people enter an elevator in a building with ten floors. What is the expected number of stopping? What assumptions do you need to calculate this expectation?

If we treat each passenger's decision about whether to stop at a certain floor as a Bernoulli trial, we could approach this question using the binomial distribution. The assumptions include:

- 8 passengers make independent decisions;
- Assume everyone enters at the ground floor, and there are 10 choices, from 1 to 10 floors. (if you assume no one stops at the first floor, then there are only 9 choices).

There are eight passengers in total, for each floor, if anyone wants to get off, the elevator will stop. Rather than calculating the probability of the elevator stops at a certain floor, we can calculate the probability of not stopping. For any floor, the probability that the elevator does not stop at any floor is:

$$\left(\frac{9}{10}\right)^8$$

the probability of the elevator stops at any floor is:

$$1 - \left(\frac{9}{10}\right)^8$$

To find the expected number of stopping in this case, define a random variable X as the number of stopping for the elevator, X follows the binomial distribution:

$$X \sim Bi\left(10, \left(1 - \left(\frac{9}{10}\right)^8\right)\right)$$

Where n=10, p=1-(9/10)^8. The expected value of the binomial distributed random variable is np:

$$E(X) = 10 * \left(1 - \left(\frac{9}{10}\right)^8\right)$$

Conditional Probability

Que 7) A person flips an unbiased coin over and over again. Player A looks for the sequence HHT and player B looks for the sequence HTT. What is the probability that player A encounters their sequence first?

The coin is unbiased so that P(H) = P(T) =0.5. Let's assume the event that we get HHT earlier than HTT is event E, and we have:

- P(E) = P(E|H)*P(H) + P(E|T)*P(T)

Where P(E|H) is getting H at the first toss. Since getting a Tail will help neither getting HHT nor HTT, thus:

- P(E) = P(E|T) = P(E|TT)=…

plug back to the previous equation:

- P(E) = 0.5 * P(E|H) + 0.5 * P(E) => P(E) = P(E|H)

so we only need to get P(E|H) to solve for P(E):

- P(E|H) = P(E|HH)*P(H) + P(E|HT)*P(T)

Whenever you get HH, you will definitely get HHT earlier than HTT since you only need to get one extra T to reach HHT, but you need to get two extra Ts to get to HTT. Thus P(E|HH) is 1:

- P(E|H) = 1*0.5 + P(E|HT)*0.5

For the same logic:

- P(E|HT) = P(E|HTH)*P(H) + P(E|HTT)*P(T)

P(E|HTT) =0 because we already reach HTT first. When seeing HTH, the first two tosses are ignorable, it is like rolling the first H. Thus:

- P(E|HTH) = P(E|H)

Now we have:

- P(E|HT) = P(E|HTH)*P(H) + P(E|HTT)*P(T)=P(E|H)*0.5+0

Plug back to the previous equation:

- P(E|H) = 1*0.5 + P(E|HT)*0.5=0.5+P(E|H)*0.5*0.5 => P(E|H)= 2/3
- P(E) = P(E|H) =2/3

Qur 8) (Part A): Mr. Jones has two children. The older child is a girl. What is the probability that both children are girls? (Part B): Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?

For a family with two children, there are four possibilities for the gender combination:

Younger	Older
G	G
B	G
G	B
B	B

four possibilities

Part A: If we know the older child is a girl, then there are only two possibilities:

Younger	Older
G	G
B	G

two possibilities
and one of them is having two girls. The probability is 1/2;

Part B: for at least one child is a boy, we have three possibilities:

Younger	Older
B	G
G	B
B	B

three possibilities
and one of them is having two boys. The probability is 1/3.

Que 9) You are given a choice of three doors by an Angel. You can choose only one of the doors among the three. Out of these three doors, two contain nothing and one has a jackpot. After you choose one of the doors, the angel reveals one of the other two doors behind which there is nothing. The angel gives you an opportunity to change the door or you can stick with your chosen door. You don't know behind which door we have nothing. Should you switch or does it not matter?

We have three doors, door 1, door 2, and door 3. Each of them has 1/3 of the chance to have the Jackpot behind. We need to decide to switch or not based on whether we have selected correctly in the first place, and compare the probability of switching and not switching. Suppose event A is selecting correctly at the first try:

- $P(A) = 1/3$; $P(\text{not } A) = 2/3$;

Let event B be winning after switching:

- $P(B|A) = 0$: if you selected correctly at the first try(event A), and you switch, you will not win anymore;

- P(B|not A) = 1: if you selected wrongly at the first try (event not A) after the angel removes another wrong door, you will definitely win after switching.

Thus:

- P(B) = P(B|A)*P(A) + P(B|not A)*P(not A) = 2/3.

Let event Not B be winning when not switching:

- P(Not B|A) = 1: if you selected correctly at the first try(event A), and you switch, you will win the Jackpot;
- P(Not B|not A) = 0: if you selected wrongly at the first try(event not A), you will lose if you choose to not switch.

Thus:

- P(Not B) = P(Not B|A)*P(A) + P(Not B|not A)*P(not A) = 1/3.

Comparing the chance of winning when switching and not switching:

- P(B) > P(Not B)

You should switch!

Bayesian Probability

Que 10) There are four boxes: A, B, C, D. John put a ball randomly in one of the four boxes and let David guess which box he put the ball. David guessed that the ball is in box A, but he was not sure. John gives him a hint that the ball is not in box B. At this time, what is the probability that the ball is in box C?

After John gave the hint, there are three situations:
- S1: A=1, B=0, C=0, D=0;
- S2: A=0, B=0, C=1, D=0;
- S3: A=0, B=0, C=0, D=1;

Define the event that John says the ball is not in box B as event B. We need to calculate the conditional probability of P(S2|B). According to Bayes Theorem:

$$P(S2|B) = \frac{P(B|S2) * P(S2)}{P(B)}$$

Let's check the elements individually:

- P(S2) is the probability that John puts the ball in box C out of the four boxes.The probability should be the same as if he puts the ball in any box. Thus:

$$P(S1) = P(S2) = P(S3) = \frac{1}{4}$$

- P(B|S2) is the conditional probability that when John puts the ball in box C, he gives the hint that the ball is not in box B when David already chose box A. After David chose A, there are only three boxes left for John to rule out, which are B, C, D. Given that the ball is actually in box C, John can only help David rule out B or D by giving him the hint. The probability of him choosing B out of the two boxes (B, D) is 1/2:

$$P(B|S2) = \frac{1}{2}$$

- According to Bayes Formula:

$$P(B) = P(B|S1) * P(S1) + P(B|S2) * P(S2) + P(B|S3) * P(S3)$$

We know that P(B|S2) is 1/2, we can use the same logic to get P(B|S1) and P(B|S3). When the ball is in box A(S1), David already got the right answer. John has three choices to give the hint: not in B, not in C, and not in D. Thus P(B|S1) is 1/3. P(B|S3) is 1/2 as John can say not in B or not in C. Combining all the information together, we will have:

$$P(B) = \frac{1}{3} * \frac{1}{4} + \frac{1}{2} * \frac{1}{4} + \frac{1}{2} * \frac{1}{4} = \frac{1}{3}$$

Thus, we have:

$$P(S2|B) = \frac{\frac{1}{2} * \frac{1}{4}}{\frac{1}{3}} = \frac{3}{8}$$

Takeaway: If John doesn't give any hint, we know the probability that the ball is in box C is 1/4. John giving the hint increases the probability of ball in box C because we are updating probability with new information, and that is the key of Bayes' theorem.

Que 11) 50% of all people who receive a first interview receive a second interview; 95% of your friends that got a second interview felt they had a good first interview; 75% of your friends that DID NOT get a second interview felt they had a good first interview. If you feel that you had a good first interview, what is the probability you will receive a second interview? (Q16 from this article)

The key to solving problems like this is to define the events carefully. Suppose your friends are a good representation of the entire population:

- Let's define feel good about the first interview as event A and define receive the second interview event B;
- "50% of all people who receive a first interview receive a second interview" means that P(B)=0.5, thus P(not B) is one minus P(B), which is 0.5 as well;
- "95% of your friends that got a second interview felt they had a good first interview" means P(A|B) =0.95;
- "75% of your friends that DID NOT get a second interview felt they had a good first interview" means P(A|not B) = 0.75.
- The question is asking given P(B), P(A|B), P(A|not B), what is P(B|A)? (If you feel that you had a good first interview, what is the probability you will receive a second interview?)

According to Bayes' theorem:

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)} = \frac{P(A|B) * P(B)}{P(A|B) * P(B) + P(A|notB) * P(notB)}$$

thus:

$$P(B|A) = \frac{0.95 * 0.5}{0.95 * 0.5 + 0.75 * 0.5} = \frac{19}{34}$$

Que 12) Suppose that in the world exist a very rare disease. The chance for anyone to have this disease is 0.1%. You want to know whether you are infected so you go take a test, and the test results come positive. The accuracy of the test is 99%, meaning that 99% of the people who have the disease will test positive, and 99% of the people who do not have the disease will test negative (Many thanks to Xavier Lavenir for correcting the assumptions in the question). What is the chance that you are actually infected? (Thanks to

Dennis Meisner for catching the error of misinterpretation here)

Assume event A is having the disease, and event B is testing positive. According to the information from the question:

- P(A) = 0.1%, then P(not A) is 99.9%;
- P(B|A) = 99%, and 1% of the people who tested positive doesn't have the disease, so that P(B|not A) =1%;
- What is P(A|B)?
  From Bayes Theorem:

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

and:

$$P(B) = P(B|A) * P(A) + P(B|notA) * P(notA)$$

plugin all the numbers:

$$P(A|B) = \frac{0.99 * 0.001}{0.99 * 0.001 + 0.01 * 0.999} = 9\%$$

Source:
- https://towardsdatascience.com/12-probability-practice-questions-for-data-science-interviews-2ec5230304d9