

## Probability Notes

The Let's Make a Deal Problem is as follows:

There is a car behind one of three doors, A, B or C, and a goat behind the other two. A contestant picks one of the three doors. The host then reveals one of the other two doors and shows a goat. The contestant is then given a chance to “stay” or “switch” doors. The question is, what is the probability of the contestant winning if they stay? Also, what is the probability if they switch?

Let's consider the strategy of staying. Your chance of initially picking the correct door (with the car) is  $1/3$ . If the strategy is staying, then you will only get the car in the cases when you picked the door correctly to begin with. Thus, in this case, your probability of winning is  $1/3$ .

Let's now consider the strategy of switching. There is a  $2/3$  chance of initially picking the incorrect door. When an incorrect door is chosen, the host is *forced* to choose the other incorrect door to reveal. Thus, if you switch, you will get the correct door. Thus, the probability of winning when you switch is  $2/3$ . (You only lose in this case if you initially pick the correct door, which only happens  $1/3$  of the time.)

This problem illustrates that our intuition about probability is often misleading, just as it is about counting. Most people feel that your probability of winning “goes up” when you are presented with the door revealing the goat. But in fact, this only occurs if you switch doors.

**In general, the probability of an event occurring is the number of successes divided by the total number of possible outcomes (known as the sample space) – assuming that each outcome is equally likely.**

**For example, given a six-sided die the probability of rolling a 2 or a 5 is  $2/6$  because there are six possibilities, of which, two are “successes.”**

**However, when rolling two dice, the probability of rolling a sum of 2 is NOT  $1/11$ . (There are 11 outcomes for the sum, 2 through 12.) This is because each of these possible sums, 2 through 12, are not equally likely.**

**The real sample space is the 36 ordered pairs  $(x,y)$  where  $1 \leq x,y \leq 6$ .**

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

**For example, the probability of rolling a 2 is  $1/36$ , since only one of these 36 possibilities adds up to 2. The probability of rolling a 3 is  $2/36$ , a 4 is  $3/36$ , etc. (Seven is the most frequent value and occurs with probability  $1/6$ .)**

**This should make sense because after you roll your first die, regardless of what you get, there will always be exactly one number you can roll on the second one to make the roll sum to 7. (This is not true of any other value. If you want an 8 but you get a 1 first, you are screwed.)**

## The Florida Lotto

Now, let's look at another problem, the Florida Lotto. There are 52 numbers to choose from (1 through 52). On a ticket you pick 6 of the numbers. If all six match, you win. There are also prizes for picking 3, 4, or 5 correctly. Let's calculate the probability of picking 3, 4, 5 or 6 of the numbers correctly.

First, let's determine the sample space. There are  $\binom{52}{6}$  possible choices of six numbers out of 52.

Of these choices, there is only 1 that corresponds to picking all six numbers correctly. Thus, the probability of winning the Florida Lotto is  $\frac{1}{\binom{52}{6}}$ .

Now, let's count the number of ways of picking 5 numbers correctly. There are  $\binom{6}{5}$  ways to choose 5 correct numbers.

These are then paired with 1 incorrect number, which can be chosen in  $\binom{46}{1}$ , since there are 46 incorrect numbers to choose from. Thus, the total number of ways to choose a combination of 5 correct numbers is  $\binom{6}{5}\binom{46}{1}$  and the probability of picking

exactly 5 numbers correct on a single ticket is  $\frac{\binom{6}{5}\binom{46}{1}}{\binom{52}{6}}$ .

In general, for  $0 \leq k \leq 6$ , the probability of picking  $k$  numbers correct is  $\frac{\binom{6}{k} \binom{46}{6-k}}{\binom{52}{6}}$ .

## Some Notation

We denote the probability of an event  $A$  as  $p(A)$ , or  $\Pr(A)$ .

The probability of an event  $A$  not happening is denoted as  $p(\bar{A})$ , or  $p(\sim A)$ .

It follows that  $p(\bar{A}) = 1 - p(A)$ .

Furthermore, all probabilities are in between 0 and 1 inclusive, and the sum of the probabilities of all possible disjoint events is always 1.

If two events  $A$  and  $B$  are mutually exclusive, then  $p(A \cap B) = 0$ . Note that  $\cap$  roughly translates to “and.” Furthermore, if two events are mutually exclusive, then  $p(A \cup B) = p(A) + p(B)$ . Note that  $\cup$  roughly translates to “or.”

If two events  $A$  and  $B$  are independent, then  $p(A \cap B) = p(A)p(B)$ . Essentially, if two events are independent, that means that knowledge of whether or not one of the events occurred does NOT affect the probability of the other event occurring. For example, the probability that it rains does not usually affect the probability that the Magic will win a basketball game. (They play indoors!)

## **Examples of Mutually Exclusive and Independent Events**

**What's the probability of picking a red Ace or a Black face card when picking one card out of a regular deck of cards?**

**Let A represent the event of getting a red Ace and B be the event of getting a black face card. Events A and B are mutually exclusive, meaning that if event A occurs, B can NOT occur.**

**$p(A) = 2/52$  because there are two red Aces.**

**$p(B) = 6/52$  because there are six black face cards (if an ace counts as a face card, then it would be  $8/52$ )**

**The probability of either occurring,  $p(A \cup B) = p(A) + p(B) = 8/52 = 2/13$ .**

**You roll a pair of fair six-sided dice and flip a fair coin. What is the probability of rolling a sum greater than 7 and getting heads.**

**Let A represent the event of rolling greater than a 7 and let B represent the event of flipping heads.**

**$p(A) = 15/36$ , (obtained by counting the appropriate ordered pairs from the previous list)**

**$p(B) = 1/2$ , since there are two sides to a coin**

**$p(A \cap B) = p(A)p(B) = 5/24$ .**

## **Inclusion-Exclusion Principle**

**Although some events are mutually exclusive and others are independent, there are pairs of events that are neither. (For example, knowing that it is going to rain affects the maximum temperature for the day.) The Inclusion-Exclusion Principle is valid for all pairs of events:**

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

**Intuitively, this says that if you want to determine the probability that either event A or event B occurs, you add up the probability of each, (but in doing that, you are double counting the events that are part of both A and B), so you subtract out the probability that both occur.**

**Consider the following problem:**

**What is the probability of picking an Ace or a Heart when choosing a single card out of a standard deck of 52 cards?**

**Let A represent the event of picking an Ace.**

**Let B represent the event of picking a Heart.**

$$p(A) = 1/13 \text{ (since there are 4 Aces out of 52 cards)}$$

$$p(B) = 1/4 \text{ (since there are 13 Hearts out of 52 cards)}$$

$$p(A \cap B) = 1/52 \text{ (since there is only one Ace of Hearts)}$$

$$\begin{aligned} p(A \cup B) &= p(A) + p(B) - p(A \cap B) \\ &= 1/13 + 1/4 - 1/52 = 4/13 \end{aligned}$$

## Conditional Probabilities

As mentioned above, sometimes probabilities change based on knowing whether or not some event has occurred.

The notation  $p(B | A)$  represents the probability that B occurs given that A has occurred. This value may very well be different than  $p(B)$ . In particular,  $p(B | A) = \frac{p(B \cap A)}{p(A)}$ .

Consider the following situation:

When it does NOT rain, Sue gets to school on time 80% of the time. When it does rain, she only gets to school on time 60% of the time.

The actual probability she gets to school on time is not either 60% or 80%, but somewhere in between. But 60% and 80% represent the probabilities she gets to school on time given that it does NOT rain or that it does rain, respectively.

Let A represent the event it rains and let B represent the event that Sue gets to school on time. Thus,  $p(B | A) = 60\%$  and  $p(B | \bar{A}) = 80\%$ .

If we add the information that it rains 40% of the time, we can now calculate the probability that Sue gets to school:

$$p(B) = p(A) \cdot p(B|A) + p(\bar{A}) \cdot p(B | \bar{A}) = .4(.6) + .6(.8) = 72\%$$

Intuitively, the probability Sue gets to school on-time can be split into two disjoint probabilities: that she gets to school on time while it raining, and that she does when it's NOT raining.

## **Tree Diagrams and Conditional Probabilities**

**Problems like this lend themselves well to a tree diagram like so:**

**In this diagram, the initial branch is a regular probability (without conditional information), but each following branch IS a conditional probability since it assumes that the previous branches have already “occurred.”**

**In general, these diagrams are helpful in situations where you get a question with some conditional probabilities and some absolute probabilities.**

**Consider the following problem:**

**There is a disease that occurs in .01% of the population. There is a test to see whether or not an individual has the disease. Given that an individual HAS the disease, the test correctly says so 99% of the time. Given that the individual does NOT have the disease, the test is correct 97% of the time. Given that an individual has tested positive for the disease, what is the ACTUAL probability that he/she has the disease?**

**First make a quick guess as to what the answer is.**

**Then, carefully work the problem out by drawing a probability tree, with the first branch indicating whether or not you have the disease, and the second set of branches indicating whether or not the TEST indicates that you have the disease. Since this is a conditional probability question, you must take the probability that you test positive AND have the disease and divide that by the probability that you test positive for the disease, since that's the given information.**