

# *Combinatorics: The Fine Art of Counting*

## Week 6 Lecture Notes – Discrete Probability

Note – Binomial coefficients are written horizontally. The symbol  $\sim$  is used to mean approximately equal.

### Introduction and Definitions

Probability is often considered a confusing topic. There are three primary sources of this confusion: (1) failure to use precise definitions, (2) improper intuition, (3) counting mistakes. Being adept counters at this point, we will hopefully avoid (3). Instances of (2) generally involve mistaken assumptions about randomness (e.g. the various “Gambler’s fallacies”) or underestimations of the likelihood of streaks or coincidences. The best way to avoid falling prey to improper intuition is to not to rely on intuition, but rather to compute probabilities carefully and rigorously. With experience, your intuition will become more trustworthy.

This leads us to addressing (1). We will begin with some definitions that may seem overly formal at first, but they will prove to be extremely useful. After you have used them to solve several problems they will become second nature. We won’t always explicitly mention all these details every time we solve a particular problem, but anytime we are uncertain about the correctness of a given approach, we will be able to fall back on our precise definitions.

**Definition:** A *sample space*  $U$  is a set whose elements are called sample points.

Sample points typically represent a complete description of an experimental outcome. For example if a coin is flipped 3 times, a particular sample point might be the sequence THH. In this case  $U$  would be the set of all such sequences – { TTT, TTH, THT, THH, HTT, HTH, HHT, HHH }

In another situation, a sample point might be a sequence of five cards selected from a deck of 52 cards, e.g. (5♠, A♦, 3♣, J♦, 5♣), and  $U$  would be the set of all such sequences –  $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48$  of them. Alternatively, if we don’t care to distinguish the order in which the cards were drawn, we might instead choose to make our sample points subsets of five cards, in which case  $U$  would be the set of all such subsets -  $\binom{52}{5}$  of them.

When solving a particular problem we often have some flexibility in defining the sample space (as in the second example above). Our choice of sample space will generally be guided by the following criteria:

1. The sample space should capture all of the information necessary to analyze the problem. If we are analyzing a situation where several different things are happening, we want each sample point to contain a record of everything that happened.
2. We want the sample space to be as simple as possible. We may choose to ignore information which is not needed to solve the problem (e.g. ignoring order), however...
3. We want the probabilities to be easy to compute. This typically means that we will want to make the sample space to be as uniform as possible. This may mean including more information than is strictly necessary to solve the problem (e.g. labeling otherwise indistinguishable objects).

**Definition:** An *event*  $A$  is a subset of  $U$ .

An event may contain one, many, all or none of the sample points in  $U$ . Getting more heads than tails in a sequence of coin flips, or getting a full-house in a poker hand are both examples of events. For finite sample spaces typically every subset of  $U$  is considered an event, and the simplest events (often called *atomic* or *elementary* events) contain a single sample point, but this need not be the case. What is required is that the collection of all events is non-empty and satisfies the following:

1. If  $A$  is an event, so is  $A^c$ , the event that  $A$  doesn't happen.
2. If  $A$  and  $B$  are events, so is  $A \cap B$ , the event that  $A$  or  $B$  happens.
3. If  $A$  and  $B$  are events, so is  $A \cup B$ , the event that  $A$  and  $B$  happen.

Note that these axioms imply that both  $U$  and the empty set are events since these are the union and intersections of an event with its complement. As a concrete example, let  $U$  be the sample space of all sequences of three coin tosses described above. Consider the following event:

$A = \{ HTT, HTH, HHT, HHH \}$	The first flip was heads.
$A^c = \{ TTT, TTH, THT, TTT \}$	The first flip was not a head.
$B = \{ TTH, THH, HTH, HHH \}$	The third coin flip is heads.
$A \cup B = \{ TTH, THH, HTT, HTH, HHT, HHH \}$	Either the first or third flip was heads.
$A \cap B = \{ HTH, HHH \}$	Both the first and third flip were heads.
$C = \{ THH, HTH, HHT, HHH \}$	There were more heads than tails.

Given a sample space and a collection of events, we are now ready to define the probability of an event.

**Definition:** A *probability measure*  $P$  is a function which assigns a real number to each event with the following properties:

1.  $0 \leq P(A) \leq 1$  for all events  $A$ .
2.  $P(\emptyset) = 0$  and  $P(U) = 1$
3.  $P(A \cap B) = P(A) + P(B)$  for *disjoint* events  $A$  and  $B$

The following lemmas are immediate consequences of the definition above:

- I.  $P(A) = 1 - P(A^c)$
- II. If  $A \subseteq B$  then  $P(A) \leq P(B)$
- III.  $P(A \cap B) = P(A) + P(B) - P(A \cup B)$  for *all* events  $A$  and  $B$

The first two follow immediately from the definitions, and the third is a consequence of the principle of inclusion/exclusion and can be generalized to more than two sets in the same way. The last definition we will need is the concept of independence.

**Definition:** Two events  $A$  and  $B$  are *independent* if and only if  $P(A \cap B) = P(A)P(B)$

This definition may seem a bit artificial at the moment, but as we shall see when we define conditional probability, this definition implies that the occurrence of event  $A$  does not change the probability of event  $B$  happening (or vice versa). This correlates with our intuitive notion of independent events, the standard example being successive coin tosses. If event  $A$  is getting a head on the first of three coin tosses and event  $B$  represents getting a head on the third coin toss, then we typically assume (with good reason) that events  $A$  and  $B$  are independent – the coin doesn't "remember" what happened on the first flip. Note that this is true whether the coin is a fair

coin or not – we don't expect the coin to change its probabilistic behavior from one toss to the next even if it is biased in general.

Note that **independent events are not disjoint** (unless one of them has probability zero). Consider the example of coin flips above where  $A \cup B = \{ HTH, HHH \}$ . The best way to think about independent events is to imagine if we restricted our sample space to consist of only events inside of A (i.e. assume that the first flip was heads and now just look at the next two flips) and adjusted our probability measure appropriately. If we now look at B inside this new sample space, it will correspond to the event  $A \cap B$  and its probability in this new space will be the same as the probability of B in the original space. Getting a head on the first toss doesn't change the probability of getting a head on the third toss. These ideas will be made more concrete when we discuss conditional probability.

There are two key things to keep in mind when computing probabilities:

**Not all events are independent** - this means we can't always multiply probabilities to determine the probability of the intersection of two events.

**Not all events are disjoint** – this means we can't always add probabilities to determine the probability of the union of two events.

These facts may both seem completely obvious at this point, but most of the mistakes made in computing probabilities amount to mistakenly assuming that two events are independent or that two events are disjoint.

In the happy situation where two events *are* independent, we can easily compute the probability of both their intersection (just multiply) and their union, since for independent events (3) above simplifies to:

$$P(A \cap B) = P(A) \cdot P(B) \quad \text{for independent events A and B}$$

## Computing Probabilities

It may seem like we haven't really gotten anywhere with all these definitions. We have to define the probability measure P for all events, but this includes the event whose probability we don't know and are trying to figure out! The key idea is to define P for a particularly simple collection of elementary events whose probability is easy to determine. We then define P for the intersection of any collection of elementary events (in most cases this will be trivial since the elementary events will either be disjoint or independent). Once we have defined P in this way, its value for any event formed using a combination of unions, intersections, and complements of elementary events can be determined by the axioms and properties of a probability measure.

The simplest example of a probability measure is the uniform probability measure defined on a finite sample space U. In this situation our elementary events will contain a single sample point and have probability  $1/|U|$ . The intersection of any two elementary events is the empty set which has probability zero. Since any event A is simply the union of some collection of  $|A|$  disjoint elementary events (one for each element of A), we can compute the probability of A by simply adding up the probability of the elementary events it contains, resulting in:

$$P(A) = |A|/|U| \quad \text{uniform probability measure on a finite sample space}$$

We have already seen many examples that fall into this category. In these cases, once the sample space has been properly specified, computing probabilities is simply a matter of counting two sets (the event we are interested in, and the whole sample space) and then computing their ratio. However if the set in question is not easy to count directly, it may be simpler to use some of the properties and axioms defined above to compute its probability (this effectively gives us another way to count the set A, assuming we know the size of U).

**Problem #1:** If a six-sided die is rolled six times, what is the probability of rolling at least one six?

**Solution #1:** Our sample space will be all sequences of six integers in the set  $[6] = \{1, 2, \dots, 6\}$  and since we assume each sequence is equally likely, we will use the uniform probability measure. If we let  $A$  be the event of rolling at least one six, and let  $B_i$  be the event of rolling a six on the  $i^{\text{th}}$  roll. We see that  $A = B_1 \cup B_2 \cup B_3 \cup B_4 \cup B_5 \cup B_6$ . We can easily compute  $P(B_i) = 1/6$ , but since the  $B_i$  are not disjoint we can't simply add their probabilities. Instead we will note that the complement of  $A$  is equal to the intersection of the complements of each of the  $B_i$ 's, i.e. the event of not rolling any sixes is equal to the event of not rolling a six on the first roll, nor the second roll, and so on. Thus  $A^c = B_1^c \cap B_2^c \cap B_3^c \cap B_4^c \cap B_5^c \cap B_6^c$ . The events  $B_i^c$  are all independent, so we can multiply their probabilities to compute the probability of  $A^c$  and then  $A$ . Putting this all together we obtain  $P(A) = 1 - (1-1/6)^6 \sim 0.665$ .

We could have defined the uniform probability measure in the example above in two different ways, either by defining  $6^6$  elementary events that each contain a single sequence and assigning them each the probability  $1/6^6$ , or by defining  $6 \times 6$  elementary events, each of which corresponds to rolling a particular number on a particular roll and has probability  $1/6$  (e.g. the third roll is a 2). In the second case, the elementary events are all either independent (if they correspond to different rolls) or disjoint (if they correspond to the same roll), so the probability of the intersection of any of them is easily computed. Both definitions result in the same probability measure since any particular sequence of rolls is the intersection of six events which specify the result of each roll.

Alternatively, we could have chosen a much smaller sample space, e.g. one with just 7 sample points each representing the number of sixes rolled and singleton elementary events for each sample point, but this would not have been very helpful because we would have to compute the probability of the problem we were trying to solve just to define our probability measure. Using a simpler probability measure on a larger sample space gives us a better starting point.

**Problem #2:** If three (not necessarily distinct) integers in  $[10] = \{1, 2, \dots, 10\}$  are randomly chosen and multiplied together, what is the probability that their product is even?

**Solution #2:** Our sample space is all sequences of three integers in  $[10]$  using the uniform probability measure. Note that this sample space is essentially the same as the one in example #1, except we have three 10-sided dice. The event we are interested in is the complement of the event that the product is odd, which is the intersection of the events that each particular number is odd. These events are independent, so we can compute the probability of their intersection by multiplying. The probability that a particular number is odd is  $5/10 = 1/2$ , so the desired probability is  $1 - (1/2)^3 = 7/8$ .

**Problem #3:** If three *distinct* integers in  $[10]$  are randomly chosen and multiplied together, what is the probability that their product is even?

**Solution #3:** Applying the same approach as in example #2, our sample space is now all sequences of three distinct integers in  $[10]$ . We can define all the same events we used in #2, but in this situation computing the intersection of the events that each digit is odd is more difficult because these events are no longer independent. We will see later how to analyze this situation using conditional probability, but for the moment we will simply count the event that all three numbers are odd directly – there are  $5 \times 4 \times 3$  sequences of distinct odd numbers between 1 and 10 and  $10 \times 9 \times 8$  sequences in all, so the desired probability is  $1 - 5 \times 4 \times 3 / (10 \times 9 \times 8) = 11/12$ .

A simpler approach to problem #2 is to notice that the product of the three numbers does not depend on the order in which they are chosen. If we ignore order we can take our sample space to be all subsets of  $[10]$  of size 3 with the uniform probability measure. The event that all three numbers are odd consist of subsets  $\{1, 3, 5, 7, 9\}$  of size 3. The desired probability is then given by  $1 - \binom{5}{3} / \binom{10}{3} = 1 - 1/12 = 11/12$ .

Problems #2 and #3 above show the distinction between sampling with or without replacement.

**Sampling with replacement** corresponds to drawing a sequence of objects out of a bag, where each object drawn is replaced before drawing the next. Examples of sampling with replacement

include rolling a die, performing multiple trials of an experiment, or even flipping a coin (think of a bag holding cards labeled heads and tails).

**Sampling without replacement** corresponds to drawing a sequence of objects out of a bag without replacing the object drawn before drawing the next. Consider a bag containing the set of objects  $\{x, y, z, \dots\}$  (these could be names, numbers, colored marbles, whatever...) we then define our sample space as the set of all possible sequences of objects that could be drawn, and then define elementary events  $X_i$  to correspond to a sequence in which the object  $x$  was drawn in the  $i^{\text{th}}$  step. The key distinction between sampling with and without replacement is that if  $j > i$ :

When **sampling with replacement**, the events  $X_i$  and  $Y_j$  are **independent**

When **sampling without replacement** the events  $X_i$  and  $Y_j$  are **not independent**.

The second fact is easy to see when  $X=Y$  since we can't draw the same object twice so  $P(X_i \cap X_i) = 0$ , but  $P(X_i)P(X_i) > 0$ . When  $X \neq Y$ , the fact that  $Y$  was not drawn on the  $i^{\text{th}}$  step (since  $X$  was) effectively increases the probability that it will be drawn on the  $j^{\text{th}}$  step. Consider the case of just two objects: when sampling without replacement there are only two possible sequences, and  $P(X_1)$  and  $P(Y_2)$  are both  $\frac{1}{2}$ , but  $P(X_1 \cap Y_2) = \frac{1}{2}$  which is not equal to  $P(X_1)P(Y_2) = \frac{1}{4}$ .

**Problem #4:** Two balls are drawn from an urn containing 5 black and 5 white balls. What is the probability that two black balls are drawn? What is the probability that two balls of different color are drawn?

**Solution #4:** If we were sampling with replacement the answers would clearly be  $\frac{1}{4}$  and  $\frac{1}{2}$ . However, the wording of the problem implies we are sampling without replacement. To analyze this problem we will consider the balls to be distinct (we can always label indistinguishable objects and then peel the labels off later if need be). Since the order in which the balls are drawn does not matter, we will ignore order and just consider drawing subsets of two balls. Our sample space is all subsets of two balls drawn from a set of 10 balls with the uniform probability measure. The probability the first event is  $\binom{5}{2} / \binom{10}{2} = 2/9$  and the probability of the second event is equal to  $\binom{5}{1} \binom{5}{1} / \binom{10}{2} = 5/9$ .

Note how the probabilities in the example above differ from sampling with replacement.

**Problem #5:** Balls are drawn from an urn containing 6 black balls and 5 white balls until all the black balls have been drawn. What is the probability that the urn is empty at this point?

**Solution #5:** This is clearly sampling without replacement. Since we don't know for sure how many balls will be drawn, we will take our sample space to be all possible sequences of drawing all 11 balls out of the urn (we could always just keep going after the last black ball was drawn) with the uniform probability measure. Each sequence corresponds to a string of 6 B's and 5 W's. The event we are interested in corresponds to the event that the last ball drawn is black, i.e. all sequences that end in B. We can easily count these and divide by the size of the sample space obtaining  $\binom{10}{5} / \binom{11}{5} = 6/11$ . A simpler approach is to note that the number of sequences that end in B is the same as the number of sequences that start with B, and the probability that the first ball drawn is black is clearly  $6/11$ .

**Problem #6:** Two cards are drawn from a standard deck of 52. What is the probability that the first card is the ace of diamonds and the second card is a spade?

**Solution #6:** This is another example of sampling without replacement. Our sample space is all sequences of two distinct cards (ordered pairs) with the uniform probability measure. The probability that the first card is the ace of diamonds is clearly  $1/52$ , and the probability that the second card is a spade is  $13/52 = 1/4$ , however these events aren't necessarily independent so we can't just multiply the probabilities together. Instead we must count the sequences which have the ace of diamonds in the first position and a spade in the second – there are 13 of these. The size of the sample space is  $52 \cdot 51$ , so the probability is  $13 / (52 \cdot 51) = 1/204$  which is not equal to  $(1/4) \cdot (1/52) = 1/208$  so the events are not independent as we suspected.

If this solution appeared a bit clumsy, it is. We will see a simpler solution using conditional probability.

**Problem #7:** Two cards are drawn from a standard deck of 52. What is the probability that the first card is an ace and the second card is a spade?

**Solution #7:** We will use the same sample space as in problem #6 above. The probability that the first card is an ace is  $4/52 = 1/13$ , and the probability that the second card is a spade is  $13/52 = 1/4$ . At this point we suspect that these events may not be independent, so we will count sequences which have an ace in the first position and a spade in the second. We have to distinguish two cases, there are 12 sequences with the ace of spades followed by a second spade and  $3 \cdot 13$  sequences with a non-spade ace followed by a spade, giving  $12 + 13 \cdot 3 = 51$ . The size of the sample space is  $52 \cdot 51$ , so the probability is  $51/(52 \cdot 51) = 1/52$ . It turns out that in this case the events were independent after all!

How can we tell when two events are independent? What is the difference between problem #6 and problem #7? To investigate this further, and to add a new tool which will simplify many of the analyses we made above, we need to define the notion of conditional probability.

## Conditional Probability

**Definition:** The conditional probability of event A given that event B has occurred is denoted by  $P(A|B)$  which is defined by the equation  $P(A|B) = P(A \cap B) / P(B)$ .

The motivation for this definition is that if we know that event B has occurred, we are effectively now working in a restricted sample space equal to B, and we only want to consider the subset of event A that lies within B.

Conditional probability has some basic properties that follow immediately from the definitions:

1.  $P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$
2.  $P(A|B) = P(A)$  if and only if A and B are independent events

The first property is often known as Bayes Theorem. This equivalence makes it easy to convert between conditional probabilities since we can rewrite it as:

$$P(A|B) = P(A)P(B|A) / P(B) \quad \text{(Bayes Theorem)}$$

The second property above gives us a way to test for independence, and is in some ways a better definition of independence than the one we gave earlier. To say that the probability of event A is the same whether event B has happened or not captures what we mean by the notion of independence.

There are two basic laws of conditional probability that are very useful in solving problem. The first is simply a generalization of property (1) above:

**Law of Successive Conditioning:**

$$P(A_1 \cap A_2 \cap A_3 \dots \cap A_n) = P(A_1) * P(A_2|A_1) * P(A_3|A_1 \cap A_2) * \dots * P(A_n|A_1 \cap A_2 \cap A_3 \dots \cap A_{n-1})$$

This statement is simply a formal way of saying that the probability of a bunch of events all happening is equal to the probability that the first event happens and then the second event and so on. Remember that events are just subsets so there is no formal notion of time order among events – when we are analyzing a sample space all possible events are laid out before us and we can choose to analyze them in whatever order and combination suits us. The law of successive conditioning does however capture what we mean when we say something like “the first card drawn is an ace and then the second card is a spade”. The wonderful thing about the law of successive conditioning is that it allows us to multiply probabilities without requiring the events involved to be independent.

To see the law of successive conditioning in action let’s return to problem #3 where we counted the number of sequences containing three distinct odd integers chosen from [10]. Alternatively we can analyze this by looking at the events  $O_1$ ,  $O_2$ , and  $O_3$  corresponding to the first second and third integers being odd and compute the intersection by successive conditioning:

$$P(O_1 \cap O_2 \cap O_3) = P(O_1) * P(O_2|O_1) * P(O_3|O_1 \cap O_2) = 5/10 * 4/9 * 3/8 = 1/12.$$

Note that we could compute  $P(O_2|O_1) = P(O_1 \cap O_2) / P(O_1) = (5*4*8)/(5*9*8)$  using the definition of conditional probability (which is what we should always do when in doubt), but in this situation it is clear what  $P(O_2|O_1)$  must be. Think of sampling without replacement, if the first number drawn is odd, only 4 of the remaining 9 numbers are odd, so  $P(O_2|O_1)$  must be 4/9, and similarly  $P(O_3|O_1 \cap O_2)$  must be 3/8 since only 3 of 8 numbers remaining are odd.

The law of successive conditioning formalizes the “counting by construction” approach that we have used to solve many combinatorial problems. The second law of conditional probability is the probabilistic analogy to counting by cases and allows us to break down complicated probability problems into smaller ones by partitioning the problems into mutually exclusive cases:

**Law of Alternatives:**

**If  $A_1, A_2, \dots, A_n$  are disjoint events whose union is  $U$  then for all events  $B$ :**

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots + P(A_n)P(B|A_n)$$

With these two tools in hand, let’s look again at problems #6 and #7 above.

**Problem #6:** Two cards are drawn from a standard deck of 52. What is the probability that the first card is the ace of diamonds and the second card is a spade?

**Solution #6b:** Let  $A$  be the event that the first card is the ace of diamonds and let  $B$  be the event that the second card is a spade.  $P(A) = 1/52$ .  $P(B|A) = 13/51$ . By the law of successive conditioning,  $P(A \cap B) = P(A)P(B|A) = (1/52)*(13/51) = 1/204$ .

**Problem #7:** Two cards are drawn from a standard deck of 52. What is the probability that the first card is an ace and the second card is a spade?

**Solution #7b:** Let  $A_1$  and  $A_2$  be the events that the first card is the ace of spades or some other ace, respectively, and let  $B$  be the event that the second card is a spade. The events  $A_1$  and  $A_2$  partition the sample space so we can apply the law of alternatives.  $P(A_1) = 1/52$ ,  $P(A_2) = 3/52$ ,  $P(B|A_1) = 12/51$ , and  $P(B|A_2) = 13/51$ .  $P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) = (1/52)*(12/51) + (3/52)*(13/51) = 1/52$ .

This solution is essentially the same as our original solution – we counted by cases there as well.

To analyze problem #7b in a more sophisticated way, let  $A = A_1 \cup A_2$  be the event that the first card is an ace. We know that  $P(A) = 1/13$  and  $P(B) = 1/4$ . We will prove that these two events are independent and therefore  $P(A \cap B) = 1/52$ .

Let  $C$  be the event that the *first* card is a spade. Clearly  $P(C) = P(B) = 1/4$ , but also note that  $P(C|A) = 1/4$  because  $1/4$  of the aces are spades, just as  $1/4$  of the deck is spades. Thus  $A$  and  $C$  are independent events. Now consider a sequence contained in the event  $A \cap C$ , i.e. any sequence that starts with the ace of spades. If we interchange the suits of the first and second card but leave the ranks intact, we obtain a new sequence of two cards which is contained in the event  $A \cap B$ . Since this process is reversible, there is a bijection between  $A \cap C$  and  $A \cap B$  so these events have the same size (as sets) and therefore the same probability, since we are using the uniform probability measure. But if  $P(B) = P(C)$  and  $P(A \cap B) = P(A \cap C)$ , then it must be the case that  $P(B|A) = P(C|A) = 1/4 = P(B)$  which means that  $A$  and  $B$  are also independent events.

The analysis above is an excellent illustration of what it means for two events to be independent. It does not mean that the events are completely unrelated, rather it means that they intersect each other in a uniform way. If we separate a deck of 52 cards into two piles, one containing the four aces and the other containing the remaining 48 cards, exactly  $1/4$  of both piles will be spades. Conversely, if we separated the deck into spade and non-spade piles, exactly  $1/13$  of both piles would be aces.

There are a lot of real world situations where conditional probability plays a critical role. Here is one example:

**Problem #8:** A certain incurable rare disease affects 1 out of every 100,000 people. There is a test for the disease which is 99% accurate. Given that you have tested positive for the disease, what is the probability that you have the disease? What if the disease affects 1 in 10 people?

**Solution #8:** We will take our sample space to be all possible sequences of two integers, where the first ranges from 1 to 100,000 and the second ranges from 1 to 100 using the uniform probability measure (imagine rolling two dice, one 100,000 sided die, and one 100 sided die). Let  $S$  be the event consisting of sequences where the first number is 1 (corresponding to having the disease), and let  $H$  be the complement of  $S$  (corresponding to being healthy). Let  $W$  be the event consisting of all sequences where the second number is 1 (corresponding to the test being wrong), and let  $R$  be the complement of  $W$  (corresponding to the test being right). Let  $A$  be the event consisting of sequences that test positive, i.e.  $A$  is the union of  $H \cap W$  and  $S \cap R$ . We want to compute  $P(S|A)$ .  $H$  and  $W$  are independent (recall how we marked the tickets) as are  $S$  and  $R$ , and  $H \cap W$  and  $S \cap R$  are disjoint, so  $P(A) = P(H) \cdot P(W) + P(S) \cdot P(R) = .99999 \cdot .01 + .00001 \cdot .99 \sim .01$ . Since  $S \cap A$  is equal to  $S \cap R$  and  $S$  and  $R$  are independent,  $P(S \cap A) = .00001 \cdot .99 \sim .0000099$ . We can now compute  $P(S|A) = P(S \cap A) / P(A) \sim .0000099 / .01 = .00099$ , which means the probability that you actually have the disease is  $\sim 1/1000$ . If instead the disease affected 1 in 10 people, we would have found  $P(A) = .9 \cdot .01 + .1 \cdot .99 = .108$  and  $P(S \cap A) = .1 \cdot .99 = .099$ , resulting in  $P(S|A) \sim .917$  or **more than 90%**.

A simple way to get a rough estimate of the probability above is to imagine a group of 100,000 people with just 1 sick person. 1000 of the group will test positive on average, and if you are one of those 1000 people, the probability that you are the 1 sick person is close to 1/1000.

Note that we could have used a different sample space above, using sequences of two different coin flips, one biased so that heads had probability 1/100,000 and the other biased so that heads had probability 1/100. This would have made the sample space smaller, but we would have needed to use a non-uniform probability measure. The probability of all the events involved would have been the same.

Before leaving the topic of conditional probability, we should mention a number of classic probability "paradoxes" that are paradoxical only to those who don't know how to compute conditional probabilities.

**Problem #9:** A family has two children. One of them is a boy. What is the probability that the other is also a boy? (assume children are boys or girls with equal probability)

**Solution #9:** Our sample space will be all possible sequences of the genders of the two children, i.e.  $\{bb, bg, gb, gg\}$ , with the uniform probability measure. Let  $B = \{bb, bg, gb\}$  be the event that one of the children is a boy,  $P(B) = 3/4$ . Let  $A = \{BB\}$  be the event that both children are boys,  $P(A) = 1/4$ . Note that  $A \cap B = A$ , so  $P(A \cap B) = 1/4$  (which incidentally means that  $A$  and  $B$  are not independent since  $P(A) \cdot P(B) = 3/16$ ). We want to compute the conditional probability of  $A$  given  $B$  which is simply  $P(A|B) = P(A \cap B) / P(B) = (1/4) / (3/4) = 1/3$ .

Note that if the problem had been worded slightly differently, the probability would have been different:

**Problem #10:** A family has two children. The older one is a boy. What is the probability that the younger one is also a boy? (assume children are boys or girls with equal probability)

**Solution #10:** Let  $B = \{bb, bg\}$  be the event that that the older child is a boy,  $P(B) = 1/2$ . Let  $A = \{bb\}$  as above, and also note that  $A \cap B = A$  as before, so  $P(A \cap B) = P(A) = 1/4$  and we have  $P(A|B) = (1/4) / (1/2) = 1/2$ .

Note that in both cases the events  $A$  and  $B$  were not independent since  $A$  was a subset of  $B$ . If we had considered instead the event  $C$  corresponding to the younger child being a boy,  $C$  would have been independent of  $B$  in the second case, but not the first. Since we computed the probability of the intersection of  $A$  and  $B$  directly in both cases, we made no assumptions about independence in either case.

The next example trips up a lot of people but just requires careful definition of the sample space.

**Problem #11:** There are three two-sided cards, one is blue on both sides, one is red on both sides, and one is red on one side and blue on the other. You are shown one side of one card selected at random and it is red. What is the probability that the other side is also red?

**Solution #11:** Label the card with two blue sides 1, and label one side “front” and the other “back”, label the card with two red sides 2 and label the sides similarly, and label the third card 3 with the red side labeled “front” and the blue side labeled “back”. Our sample space will be all sequences of two values, the first a number from 1 to 3 indicating the card chosen, and the second an f or a b indicating the side shown. Let A be the event that a red side is shown, then  $A = \{2F, 2B, 3F\}$  and  $P(A) = 3/6 = 1/2$ . Let B be the event that card 2 is chosen, then  $B = \{2F, 2B\}$ . We wish to compute  $P(B|A) = P(A \cap B) / P(A) = (1/3) / (1/2) = 2/3$ .

We will leave the most famous conditional probability “paradox” of all, the “Monty Hall” problem, for you to enjoy on the homework.