Probability

Probability is the study of uncertain events or outcomes. Games of chance that involve rolling dice or dealing cards are one obvious area of application. However, probability models underlie the study of many other phenomena. The number of vehicles passing through an intersection in a ten minute period, the number of people who develop a disease (and then the number that recover from the disease), the length of time it takes to search the internet on a topic, the drop in blood pressure after taking an anti-hypertensive drug, the change in the stock market index over the course of a day, etc., etc. are all examples where probability models can be applied.

In each of these situations, the result is uncertain. We will not actually know the result until after we collect the data. In all of these situations, we can use probability models to describe the patterns of variation we are likely to see.

Experiments, Trials, Outcomes, Sample Spaces

Consider some phenomenon or process that is repeatable. We will use the word *experiment* to describe the process, and we will call a single repetition of the experiment a *trial.* The data is often numerical such as the number on the top face of a die, or the blood pressure of a patient. However, data can also be *categorical* such as what comes up when a coin is tossed, or the colour that litmus paper turns when placed in an unknown solution. We call the result of the experiment an *outcome.*

The main features of these experiments is that there is more than one possible outcome, and the experiments are repeatable. Also, when we repeat the experiment we will not necessarily get the same outcome.

If we record the outcomes of several repetitions of an experiment, we obtain a set of data. The analysis of data we have already collected is dealt with in the subject of *statistics.* Predicting the patterns of variability in the outcomes of experiments is dealt with in the subject of *probability.*

We define the *sample space* as the set of all possible outcomes of an experiment.

Example 1 Our experiment consists of tossing a fair coin and observing the top face, then $S = \{H, T\}$

Our experiment consists of tossing a fair coin twice and observing the top face on each trial, then $S = \{HH, HT, TH, TT\}$

Our experiment consists of tossing a fair coin twice and counting the number of Heads, then $S = \{0, 1, 2\}$.

Example 2 Our experiment consists of counting the number of cars passing through the intersection of King St. and University Ave. between 4 and 5 pm on a Wednesday evening. Then $S = \{0, 1, 2, 3, ...\}$. Note that some of these outcomes will have an extremely low chance of occurring. Some might be physically impossible. However, this is a convenient way to think of this sample space.

Example 3 Our experiment involves measuring the weight of a randomly chosen student at the University of Waterloo. Here S is a subset of the real numbers.

Discrete and Continuous Sample Spaces

We say a sample space is *discrete* if it contains a finite or a countably infinite number of outcomes. The sample spaces in the first two examples are discrete sample spaces.

We say a sample space is *continuous* if it contains a non-countably infinite number of possible outcomes. (In reality our measuring instruments only measure to a certain number of decimal points, so all sample spaces could be considered discrete. However, it is very convenient to allow for continuous sample spaces since we can then use continuous mathematical models for descriptions of the pattern of variation.)

Events Defined on Sample Spaces

We define an *event* as a subset of a sample space. Another way to think of this is that an event is a collection of outcomes that share some property.

Example 4 A die is rolled. Define the event A to be that we observe and odd number. Then $S = \{1, 2, 3, 4, 5, 6\}$, and $A = \{1, 3, 5\}$.

Example 5 A die is rolled. Define the event B to be that we observe a number bigger than 3. Then $B = \{4, 5, 6\}$.

We define the *union* of two events as the outcomes that are in one event or the other event or both. We write $A \cup B$ and say "A or B".
We define the *intersection* of two events as the outcomes that are in both events. We write $A \cap B$, or, more simply, AB, and say A and B".

Example 6 Using the events A and B defined above, $A \cup B = \{1, 3, 4, 5, 6\}$ (odd **or** bigger than 3 **or both**), and $AB = \{5\}$ (odd **and** bigger than 3).

Probability Distributions for Discrete Sample Spaces

We often call the individual possible outcomes of an experiment *points*. With a discrete sample space, we can arrange the outcomes (points) so that one of them is the first, another is the second, etc. So let the points in the sample space be labeled $1, 2, 3, \ldots$. We assign to each point i of S a real number p_i which is called the probability of the outcome labeled i. These probabilities must be non-negative and sum to one. So the p_i's must satisfy:

$$p_i \geq 0, \text{ and } p_1 + p_2 + p_3 + \ldots = 1$$

Any set of p_i's that satisfies these results is a *probability distribution* defined on the discrete sample space.
Of course, to be useful in practice, we want a probability distribution that reflects the experimental situation. In games of chance, we often refer to *fair* dice, *fair* coins, *random* draws,

or *well-shuffled* decks of cards to indicate that .

Example 7 A *fair* die is rolled and we note the number on the top face. Then $S = \{1, 2, 3, 4, 5, 6\}$. Since the die is fair, the probability distribution $p_1 = 1/6, p_2 = 1/6, \ldots, p_6 = 1/6$ will describe the situation.

Example 8 Suppose the die was weighted so that odd numbers occurred twice as often as even numbers. Then a reasonable probability distribution would be $p_1 = 2/9, p_2 = 1/9, p_3 = 2/9, p_4 = 1/9, p_5 = 2/9, p_6 = 1/9$. Note that these sum to one and have the desired property.

Example 9 J. G. Kalbfleisch in his book *Probability and Statistical Inference, Volume 1: Probability* gives an example of a brick-shaped die. Suppose a die is brick-shaped, with the numbers 1 and 6 on the square faces and the numbers 2,3,4,5 on the rectangular faces. The square faces are $2cm \times 2cm$, and the rectangular faces are $2cm \times (2 + k)cm$, where $-2 < k < \infty$. Then we have $p_1 = p_6$, and $p_2 = p_3 = p_4 = p_5$. If $p_1 = p_6 = p$ and $p_2 = p_3 = p_4 = p_5 = q$, then $2p + 4q = 1$, so $p = 1/2 - 2q$.
Now these probabilities will be a function of k. If $k = 0$, then $p = q = 1/6$. As k approaches -2, q approaches 0. As k gets bigger, p approaches 0.
We could rewrite these probabilities as $q = 1/6 + \theta$ and $p = 1/6 - 2\theta$, where θ is a function of k. By creating several dice with different values of k, we could experiment to try to determine the relationship between θ and k. Alternatively, we could try to use the laws of physics and develop a mathematical model to relate θ and k.

Probabilities of Events

Once we have defined a probability distribution for the individual outcomes, we define the probability of any event defined on the sample space as the sum of the probabilities assigned to the outcomes in that event.

Example 10 If A is the event that we roll an odd number when a die is rolled, the probability of event A is $P(A) = p_1 + p_3 + p_5$. From the three previous examples, this would be 1/2 if the die were fair, 2/3 if it were weighted towards the odd numbers as described, and 1/2 if it were brick shaped, regardless of k.

Example 11 We can find $P(A \cup B)$ in a couple of ways. First, we can add the probabilities of the outcomes in A or B or both. So if A is the event that we get an odd number, and B is the event we get a number greater than 3 when we roll a die, $P(A \cup B) = p_1 + p_3 + p_4 + p_5 + p_6$. For the probability distributions above, this will be 5/6 or 8/9 or $5/6 - \theta$.
Note though, that if we take the outcomes in A and add to them the outcomes in B, we have all the outcomes in $A \cup B$, except the outcomes in $A \cap B$ are there twice. This suggests that $P(A \cup B) = P(A) + P(B) - P(AB)$. Here $P(A \cup B) = 1/2 + 1/2 - 1/6 = 5/6$, for the case of the fair die. We can verify the other cases.

Conditional Probability

It is very common in many applications to want to know the probability of an event *conditional on* the occurrence of some other event. For example, hypertension (high blood pressure) is very common the population. Being overweight is also common. If we sample someone from the population and see he is overweight, what is the probability that he has high blood pressure as well.

Here let A be the event that an individual has high blood pressure and B be the event that an individual is overweight. Then the probability that someone has high blood pressure *given that* he is overweight is written $P(A|B)$.

To calculate this probability, we note that we only want to deal with the overweight people. So our "sample space" is confined only to those people (in the set B). Of those overweight people, we only want the hypertensive ones. So they must be overweight *and* hypertensive (so in the set AB).

So we have $P(A|B) = P(AB)/P(B)$. This also gives us a useful way to calculate $P(AB)$, since $P(AB) = P(A|B)P(B)$

Note that we could just as well compute $P(AB) = P(B|A)P(A)$. If we put these together, we get a very useful theorem named after Thomas Bayes ($\sim$ 1702 - 1761).

Bayes' Theorem Let A and B be events defined on the same sample space. Then $P(A|B) = P(B|A)P(A)/P(B)$.

Example 12 For the events A (we roll an odd number) and B (we roll a number bigger than 3), we have the probability that we roll an odd number given that we roll a number bigger than 3 is $P(A|B) = P(AB)/P(B) = 1/3$, for the equal probability case. This makes sense since there are three numbers bigger than 3 and only one of them is odd. Note that if we use the case where the odds are weighted more heavily, $P(A|B) = (2/9)/(4/9) = 1/2$.

Independent Events

We define events A and B to be independent if and only if $P(AB) = P(A)P(B)$. Note that if events are independent, $P(A|B) = P(AB)/(PB) = P(A)P(B)/P(B) = P(A)$. This makes intuitive sense since it says that if the events are independent, knowing that B has occurred does not change the probability that A also occurred.

Example 12 Are the events A and B for the die independent? Here for the equal outcome case, $P(AB) = 1/6 \neq P(A)P(B) = 1/2 * 1/2 = 1/4$. So A and B are not independent. Knowing you have a number greater than 3 decreases your chances of getting an odd number.

A Bit of Counting

There are many situations where each of the outcomes in a sample space is equally likely. In these situations, the probability of an event A can be calculated by dividing the number of outcomes in A by the total number of outcomes in S. To calculate the total number of outcomes, we need a bit of combinatorics.

If we have 3 objects that are all different, we can arrange them in a row in 6 ways. For example, if the objects are denoted by the letters X, Y, Z, then there 6 ways to arrange

them:
$XYZ, XZY, YXZ, YZX, ZXY, ZYX$. We have 3 choices for the first letter, for each of those we have 2 choices for the second letter, and then there is only one choice left for the third letter. So there are $3 \times 2 \times 1 = 6$ different ways they can be arranged.

If there are n distinct objects, we can arrange them in $n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$ ways. We define $n!$ as $n! = n \times (n-1) \times (n-2) \times \ldots \times 2 \times 1$, to represent this number.

In many applications, we need to know the number of ways to select a subset of outcomes from the entire set. For example, we might want to select 2 items from 4 where we do not really care about the order we select them (for example, XY and YX are treated the same way). If we have letters W, X, Y, Z, there are 6 subsets of size 2: WX, WY, WZ, XY, XZ, YZ. To count these, we could say that there are 4 choices for the first letter, and three for the second. But that counts XY and YX differently. There are 2! ways to arrange the two letters we chose, so each selected pair is counted twice, so, overall, there are $\frac{4\times 3}{2!} = 6$ ways to choose 2 letters from 4 if the order does not matter. We define a special notation for this, and say "4 choose 2":

$$\binom{4}{2} = \frac{4!}{2!2!}$$

is the number of ways to choose 2 items from 4 if order does not matter. In general, the number of ways to choose x items from n if order does not matter is:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}.$$

We will need these simple counting methods to calculate probabilities of various events.

Example 13 In Lotto 6-49, you choose 6 numbers from 49. A random set of 6 numbers is chosen on the night of the draw. To calculate the probability that your numbers will be drawn, we first note that the sample space consists of $\binom{49}{6}$ possible ways to choose 6 numbers from 49. Because the numbers are chosen at random, each of the $\binom{49}{6} = 13,983,816$ possible combinations is equally likely. Hence the probability that your numbers will be drawn is:

$$\mathrm{P(win)} = \frac{1}{\binom{49}{6}} = \frac{1}{13,983,816} = .00000007151124$$

Example 14 The various other prizes from Lotto 6/49 are taken from the Lotto 6/49 website and given below. Can you verify them? (Some are quite tricky).

Match	Prize	Probability
6 of 6	Share of 80.50% of Pools Fund	1/13,983,816
5 of 6 + Bonus	Share of 5.75% of Pools Fund	1/2,330,636
5 of 6 (no bonus)	Share of 4.75% of Pools Fund	1/55,491
4 of 6	Share of 9.00% of Pools Fund	1/1,032
3 of 6	$10	1/57
2 of 6 + Bonus	$5	1/81

Independent Experiments

If we conduct two experiments so that the outcome of the second experiment is not influenced by the outcome of the first experiment, we say that the experiments are *independent*. For example, experiment 1 could consist of tossing a fair coin, and experiment 2 could consist of rolling a fair dice independently of the results of experiment 1. Then the sample space for the combined experiment (experiment 1 followed by experiment 2) is:
$S = \{(H,1),(H,2),(H,3),(H,4),(H,5),(H,6),(T,1),(T,2),(T,3),(T,4),(T,5),(T,6)\}$. Since the coin and die are fair, then the probability of a Head on the coin is 1/2, and the probability of a 1 on the die is 1/6. All 12 events in this sample space should be equally likely since the result of the second experiment did not depend on the result of the first. Hence, the probability of the combined outcomes (H,1) is 1/12 = 1/2*1/6.
In general, if p_i is the probability of outcome i in experiment 1, and q_j is the probability of outcome j in experiment 2, and the experiments are independent, then $p_i * q_j$ is the probability observing of outcome i in experiment 1 followed by outcome j in experiment 2.
Intuitively this makes sense, since in a fraction p_i of the time we will observe outcome i in experiment 1, and *of those times we observe outcome* i, we will observe outcome j in experiment 2, a fraction q_j of the time.

Independent Repetitions

A special case of independent experiments occurs when experiment 2 is an *independent repetition* of the experiment 1. Then, unless things have changed somehow, the probability distribution for the outcomes will be the same for each repetition. So the probability of outcome i on repetition 1 and outcome j on repetition 2 is $p_i * p_j$.
Being able to do independent repetitions of experiments is a very important part of the scientific method.

Example 11 A fair coin is tossed 4 times independently. Since there are 2 possibilities for each toss (H or T), there will be 2^4 possible outcomes. Since the coin is fair, each of these will be equally likely, so the probability distribution will assign probabilities 1/16 to each point in the combined sample space.

Example 13 A fair coin is tossed, independently, until a head appears. Then the sample space is $S = \{H,(T,H),(T,T,H),(T,T,T,H),\ldots\}$.

Because the tosses are independent, the probabilities assigned to the points of S are:
$\{1/2, 1/4, 1/8, 1/16, \ldots\}$. This is our first example of a finite sample space with an infinite number of outcomes. But can a sum of an infinite number of real numbers equal 1?

Our sum is $T = 1/2 + 1/4 + 1/8 + 1/16 + \ldots$. This is an example of an *infinite geometric series*, where each term after the first in the sum is 1/2 of the term before it. There is a formula for finding the sums of such series. If a is the first term in the sum, and r is the common ratio between all terms in the sum, then $T = a/(1-r)$, providing $-1 < r < 1$. This works here since $a = 1/2$, and $r = 1/2$, so $T = 1$.

Some Discrete Probability Models

Bernoulli Trials - The Binomial Distribution

A common application occurs when we have *Bernoulli Trials* - independent repetitions of an experiment in which there are only two possible outcomes. Label the outcomes as a "Success" or a "Failure". If we have n independent repetitions of an experiment, and $P(Success) = p$ on each repetition, then:

$$P(x\ Successes\ in\ n\ repetitions) = \binom{n}{x} p^x (1-p)^{(n-x)}$$

Example 15 A multiple choice exam consists of 25 questions with 5 responses to each question. If a student guesses randomly for each question, what is the probability he will answer 13 questions correctly? at least 13 questions correctly?
This is a Bernoulli Trials situation with $P(Success) = 1/5$.

Hence, $P(13\ correct) = \binom{25}{13}(1/5)^{13}(4/5)^{12} = .00029$.

The probability of at least 13 correct, is:
$P(13\ correct) + P(14\ correct) + \ldots + P(25\ correct) = 0.00037$.

Negative Binomial Distribution

In a sequence of Bernoulli Trials, suppose we want to know the probability of having to wait x trials before seeing our kth success. If the kth success occurs on the xth trial, we must have $k-1$ successes in the first $x-1$ trials and then a success on the xth trial. So the probability is:

$$P(kth\ Success\ on\ xthtrial) = \binom{x-1}{k-1} p^{k-1}(1-p)^{x-k} * p = \binom{x-1}{k-1} p^k (1-p)^{x-k}$$

Note that there is no upper limit on x, so our sample space here is $S = \{k, k+1, k+2, \ldots\}$.

Expected Values In the above examples, we are counting the number of successes in n trials or the number of trials until some event occurs. It is natural to want to know, *on average* how many successes will we see, or *on average* how many trials until some event occurs. *On average* is meant here to be a theoretical concept referring to repeating the process over and over again and averaging the numbers from an *infinite* number of trials. We refer to this theoretical average as the *Expected Value* for the number of successes or the number of trials.

If p_i represents the probability of i successes, or i trials until some event, then the *Expected Number* of successes or trials is $0 \times p_0 + 1 \times p_1 + 2 \times p_2 + \ldots + p_n$ where we are counting the number of successes. If we are counting the number of trials, we get a sum that goes to infinity.

Example 16 If you buy one ticket on Lotto 6/49 per week, how many weeks *on average* will you have to wait until you win the big prize?

From the earlier example, the probability of winning the big prize on any draw is:

$$p = \frac{1}{49\ choose 6} = \frac{1}{13,983,816}$$

So the probability that you have to wait x weeks to win the big prize is:

$$P(wait\ x\ weeks) = (1-p)^{x-1}p$$

Thus, the expected number of weeks to wait is $E = 1 \times p + 2 \times (1-p)p + 3 \times (1-p)^2 p + \ldots$. This is a combined arithmetic-geometric sequence. There are several ways to sum this. If you consider $E - (1-p)E$, you can verify that $E = 1/p$. So you need to wait 13,983,816 weeks on average to win the big prize.

Example 17 - Montmort Letter Problem Suppose n letters are distributed at random into n envelopes so that one letter goes into each envelope. On average, how many letters will end up in their correct envelopes? Does this depend on n?

There are two ways to do this question - the hard way and the easy way. Figuring out the probability that x letters end up in the correct envelopes is not that easy. So lets try this another way. Define X_i to be 0 if letter i does not end up in the correct envelope and define X_i to be 1 if letter i does end up in the correct envelope. Then $T = \sum X_i$ is the number of letters in the correct envelopes. Then the expected value of T is the sum of the expected values of the $X_i's$. Now $P(X_i = 1) = 1/n$, since there are n choices for letter i. So the expected value of X_i is also $1/n$, and so the expected value of T is $n \times 1/n = 1$. So, on average, regardless of n, we expect to find only one letter in the correct envelope.

Reference

Kalbflesich, J.G., *Probability and Statistical Inference, Volume 1: Probability, Second Edition*, Springer-Verlag, New York, 1985.