The probability axioms are the basic rules of probability theory. And they are surprisingly few. But they imply many interesting properties that we will now explore. First we will see that what you might think of as missing axioms are actually implied by the axioms already in place.

For example, we have an axiom that probabilities are non-negative. We will show that probabilities are also less than or equal to 1. We have another axiom that says that the probability of the entire sample space is 1. We will show a counterpart that the probability of the empty set is equal to 0. This makes perfect sense. The empty set has no elements, so it is impossible. There is 0 probability that the outcome of the experiment would lie in the empty set.

We also have another intuitive property. The probability that an event happens plus the probability that the vendor does not happen exhaust all possibilities. And these two probabilities together should add to 1. For instance, if the probability of heads is 0.6, then the probability of tails should be 0.4.

Finally, we can generalize the additivity axiom, which was originally given for the case of two disjoint events to the case where we're dealing with the union of several disjoint events. By disjoint here we mean that the intersection of any two of these events is the empty set. We will prove this for the case of three events and then the argument generalizes for the case where we're taking the union of k disjoint events, where k is any finite number.

So the intuition of this result is the same as for the case of two events. But we will derive it formally and we will also use it to come up with a way of calculating the probability of a finite set by simply adding the probabilities of its individual elements.

All of these statements that we just presented are intuitive. And you do not to really need to be convinced about their validity. Nevertheless, it is instructive to see how these statements follow from the axioms that we have put in place.

So we will now present the arguments based only on the three axioms that we have available. And in order to be able to refer to these axioms, let us give them some names, call them axioms A, B, and C.

We start as follows. Let us look at the sample space and a subset of that sample space. Call it A. And consider the complement of that subset. The complement is the set of all elements that do not belong to the set A. So a set together with its complement make up everything, which is the entire sample space. On the other hand, if an element belongs to a set A, it does not belong to its complement. So the intersection of a set with its complement is the empty set.

Now we argue as follows. We have that the probability of the entire sample space is equal to 1. This is true by our second axiom. Now the sample space, as we just discussed, can be written as the union of an event and the complement of that event. This is just a set theoretic relation. And next since a set and its complement our disjoint, this means that we can apply the additivity axiom and write this probability as the sum of the probability of event A with the probability of the complement of A. This is one of the relations that we had claimed and which we have now established.

Based on this relation, we can also write that the probability of an event A is equal to 1 minus the probability of the complement of that event. And because, by the non-negativity axiom this quantity here is non-negative, 1 minus something non-negative is less than or equal to 1. We're using here the non-negativity axiom. And we have established another property, namely that probabilities are always less than or equal to 1.

Finally, let us note that 1 is the probability, always, of a set plus the probability of a complement of that set. And let us use this property for the case where the set of interest is the entire sample space. Now, the probability of the entire sample space is itself equal to 1. And what is the complement of the entire sample space?

The complement of the entire sample space consists of all elements that do not belong to the sample space. But since the sample space is supposed to contain all possible elements, its complement is just the empty set. And from this relation we get the implication that the probability of the empty set is equal to 0. This establishes yet one more of the properties that we had just claimed a little earlier.

We finally come to the proof of the generalization of our additivity axiom from the case of two disjoint events to the case of three disjoint events. So we have our sample space. And within that sample space we have three events, three subsets. And these subsets are disjoint in the sense that any two of those subsets have no elements in common. And we're interested in the probability of the union of A, B, and C.

How do we make progress? We have an additivity axiom in our hands, which applies to the case of the union of two disjoint sets. Here we have three of them. But we can do the following trick. We can think of the union of A, B, and C as consisting of the union of this blue set with that green set. Formally, what we're doing is that we're expressing the union of these three sets as follows. We form one set by taking the union of A with B. And we have the other set C. And the overall union can be thought of as the union of these two sets.

Now since the three sets are disjoint, this implies that the blue set is disjoint from the green set and so we can use the additivity axiom here to write this probability as the probability of A union B plus the probability of C. And now we can use the additivity axiom once more since the sets A and B are disjoint to write the first term as probability of A plus probability of B. We carry over the last term and we have the relation that we wanted to prove.

This is the proof for the case of three events. You should be able to follow this line of proof to write an argument for the case of four events and so on. And you might want to continue by induction. And eventually you should be able to prove that if the sets A1 up to Ak are disjoint then the probability of the union of those sets is going to be equal to the sum of their individual probabilities. So this is the generalization to the case where we're dealing with the union of finitely many disjoint events.

A very useful application of this comes in the case where we want to calculate the probability of a finite set. So here we have a sample space. And within that sample space we have some particular elements S1, S2, up to Sk, k of them. And these elements together form a finite set.

What can we say about the probability of this finite set? The idea is to take this finite set that consists of k elements and think of it as the union of several little sets that contain one element each. So set theoretically what we're doing is that we're taking this set with k elements and we write it as the union of a set that contains just S1, a set that contains just the second element S2, and so on, up to the k-th element.

We're assuming, of course, that these elements are all different from each other. So in that case, these sets, these single element sets, are all disjoint. So using the additivity property for a union of k disjoint sets, we can write this as the sum of the probabilities of the different single element sets.

At this point, it is usual to start abusing, or rather, simplifying notation a little bit. Probabilities are assigned to sets. So here we're talking about the probability of a set that contains a single element. But intuitively, we can also talk as just the probability of that particular element and use this simpler notation. So when using the simpler notation, we will be talking about the probabilities of individual elements. Although in terms of formal mathematics, what we really mean is the probability of this event that's comprised only of a particular element S1 and so on.