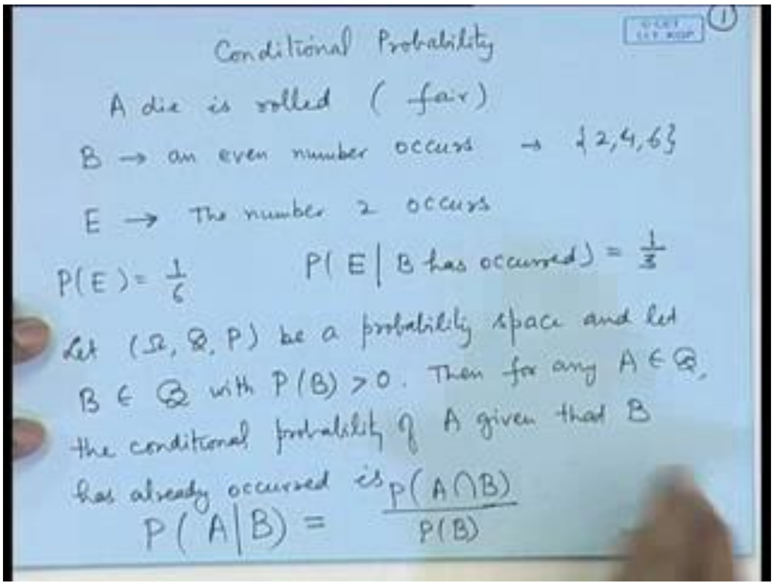
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 09
Conditional Probability

Today we are going to introduce the concept of conditional probability. So, sometimes after conducting A random experiment or after observing certain random experiment, we have some partial information about the outcome of the experiment. Now in the light of that partial information about the random experiment or the outcome, if we want to find out the probability of certain event then the probability gets modified.

(Refer Slide Time: 00:56)

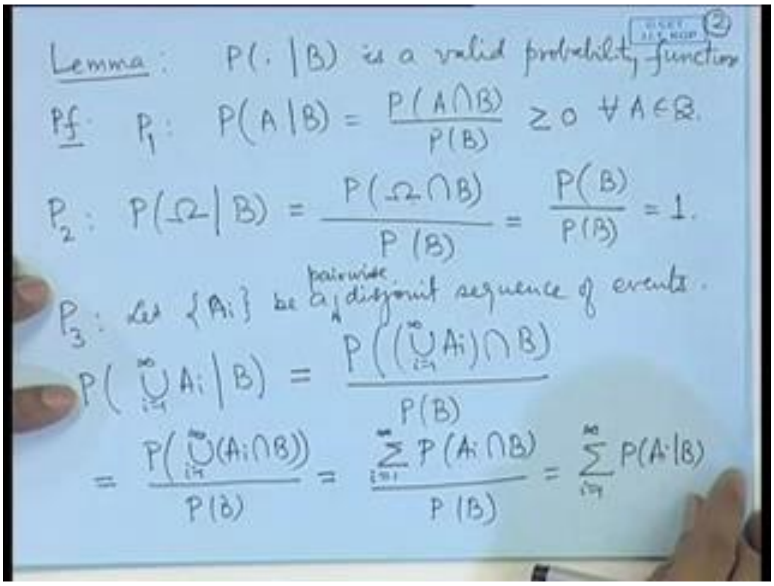


Let us take a very simple example suppose we say A die is rolled suppose it is a fair die and let say B is the event that an even number occurs and if I say E is the event that the number 2 occurs.

Now, if I want to find out probability of E it is equal to 1 by 6. However, if I know that an even number has occurred then probability of E given that B has occurred is 1 by 3 because the B has elements 2 4 and 6; that means, my sample space has been restricted to consist of 3 elements: 2 4 and 6. Now out of this if we say two occurs, then it is one of the 3 possibilities, and therefore the probability of that is equal to 1 by 3. So, you can see that the effect of the partial information makes us to make helps us in making better probability statements. So, let us define formally the conditional probability

So, if I say that omega B P be a probability space and let B be an event with positive probability then for any event A the conditional probability of A given that B has already occurred is defined by probability of A given B. So, this is the notation, probability of A given B that is defined as probability of A intersection B divided by probability of B. So, now, given any event B if we are defining the probability of A, then first of all we should check whether it is A proper probability function satisfying the axioms of the probability let us check that.

(Refer Slide Time: 04:08)

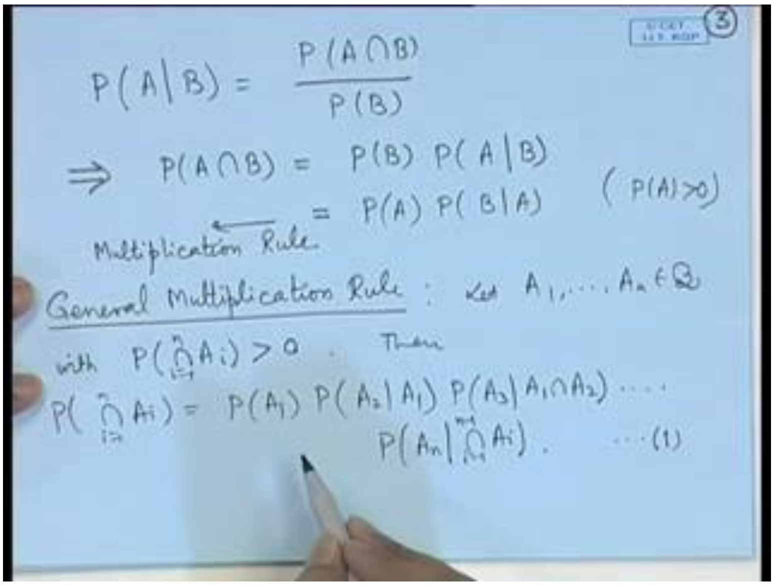


So, we state it as a result that P dot B is A valid probability function. Meaning thereby that it will satisfy the 3 axioms that is probability of A given B that is equal to probability of A intersection B divided by probability of P of B. Now according to the assumptions of the probability; probability of A intersection B is always greater than or equal to 0 probability of B is strictly positive, therefore this is always greater than or equal to 0 for all events B. So, the first axiom is satisfied.

Let us look at the second axiom probability of omega given B, this is equal to probability of omega intersection B divided by probability of B. Now note that B is a subset of omega therefore, the numerator will be probability of B divided by probability of B and therefore it is equal to 1. The third axiom is the countable addititive axiom. So, let us consider A i be a disjoint sequence of events; let us consider probability of union A i given B. Now by the definition of the conditional probability, this will be equal to probability of union A i intersection B divided by probability of B. Now in the numerator I can apply the distributive property of the unions and intersection, so this becomes probability of union of A i intersection B divided by probability of B.

Now, here note that A i s were disjoint basically we can consider them to be pair wise disjoint. So, if they are disjoint then A i intersection Bs will also be pair wise disjoint and therefore, by the axioms of the probabilities by the third axiom, I will have probability of union A i intersection B is equal to sigma i equal to 1 to infinity probability of A i intersection B divided by probability of B; where each of the term then becomes probability of A i given B. So, you can note here that probability of the union is equal to sum of the probabilities, which is the countable additivity axiom therefore; the conditional probability function is a well defined probability function. Now let us look at some direct consequences of this definition of the conditional probability.

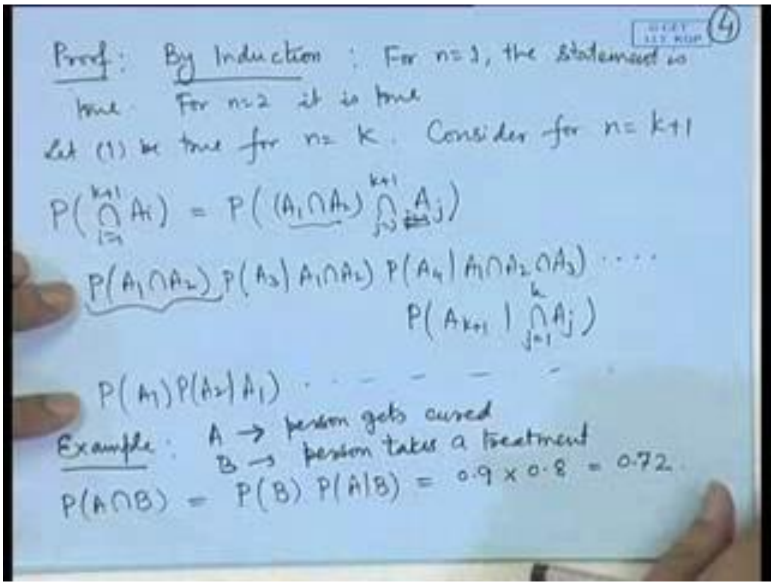
(Refer Slide Time: 07:28)



Let us consider say by statement probability of A given B is equal to probability of A intersection B divided by probability of B. Now this statement you can write as probability of A intersection B is equal to probability of B into probability of A given B. So, an interpretation of this statement is that, the probability of simultaneous occurrence of A and B is equal to the probable conditional probability of A given B multiplied by the probability of B alone happening, if we interchange the roles of A and B then we can also write it as probability of A into probability of B given A of course, provided we consider that probability of A is positive.

Now, this representation of giving the expression for the simultaneous occurrence probability in terms of a product, where in the product one term is a conditional probability and another term is a marginal probability is known as multiplication rule. So, this is called multiplication rule. So, the general multiplication rule is; let us consider events A 1, A 2 A n be events with now in order to define the conditional probabilities the probabilities must be positive. So, we can take the smallest event which may occur in the definition and we can consider the probability of that to be positive and then the probability of intersection A i is equal to 1to n can be expressed as probability of A 1 into probability of A 2 given A 1 into probability of A 3 given A 1 intersection A 2 and so on probability of A n given intersection A i as equal to 1 to n minus 1.

(Refer Slide Time: 10:22)



Let me give this statement number 1; so in order to prove this one we can follow induction. So, by induction if we want to see for n is equal to 1, the statement will reduce to simply probability of A 1 is equal to probability of A 1 which is trivially true. So, for n is equal to 1 the statement is true; for n is equal to 2 it is true as we mentioned by the definition of the conditional probability that probability of A 1 intersection A 2 will become equal to probability of A 1 into probability of A 2 given A 1; now the statement for n is equal to 2 will be used for extending from k to k plus 1.

So, let 1 be true for n is equal to K and consider for n is equal to k plus 1. So, probability of intersection A i, i is equal to 1 to k plus 1. Now this one we can express as probability of A 1 intersection A 2 intersection A j. So, j is equal to 3 to k plus 1. So, here if you see in the intersection these are k minus 1 sets and this is one set. So, this is total k sets and we have made the assumption that the statement is true for k therefore, you can write as probability of A 1 intersection A 2, into probability of A 3 given A 1 intersection A 2, into probability of A 4 given A 1 intersection A 2, intersection A 3 and so on to probability A k plus 1 given intersection A j, j is equal to 1 to k and now the first statement we can express as probability of A 1 into probability of A 2 given A 1 and all the remaining terms. Hence the statement one is true for n is equal to k plus 1 also and therefore by the principle of mathematical induction the general multiplication rule is true.

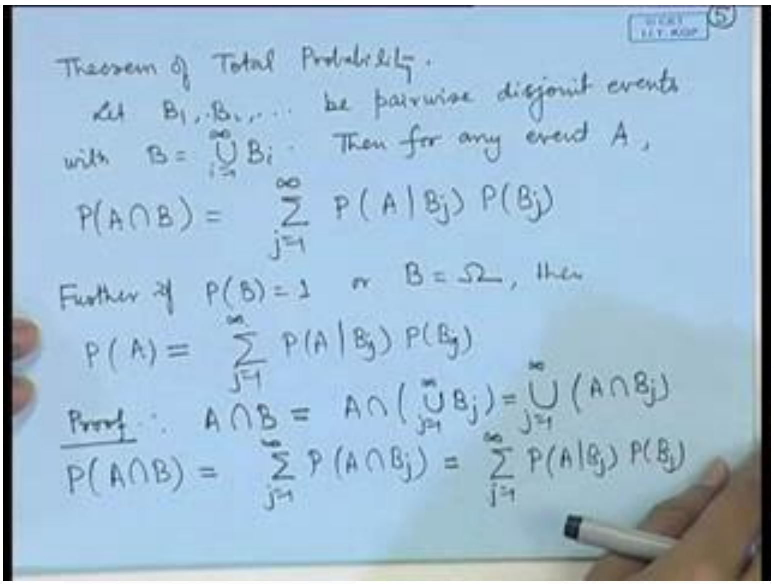
The significance of this result lies in that many times we are not aware of the probability of the simultaneous, it may be somewhat difficult because what may happen that the events are conditional that is after occurrence of one something may happen for example, if a person is sick with a certain disease and he takes a treatment, what is the probability that he will be cured? Now the probability of simultaneous occurrence of this event may not be so simple; however, if we know that the percentage of people who get cured by taking that particular treatment and how many persons eventually get recovered then this probability can be calculated.

That means if I say A is the probability of getting cured; B is suppose A is A event of getting cured and B is the event of taking A particular treatment. Now if you look at the percentage of people actually getting cured and the percentage of people who take the treatments and they are cured or in the reverse way if you look at A percentage of people who take the treatment and the out of the percentage of the people who take the treatment, how many get cured then this probability can be evaluated let me write it as a example.

Let A be the event person gets cured; B is person takes a treatment. So, now, if I look at suppose probability of A intersection B is to be determined, then we can write as probability of B into probability of A given B. Suppose we know that how many people take that particular treatment, suppose 90 percent of the people take the treatment an of the number of people who take the treatment suppose 80 percent get recovered, then this probability turns out to be 0.72, which may be might have been difficult otherwise.

A now this representation that a certain event can be considered as a consequence of certain other event or happening after a certain event, this lead to some interesting possibilities and a one of the major things is so called theorem of total probability.

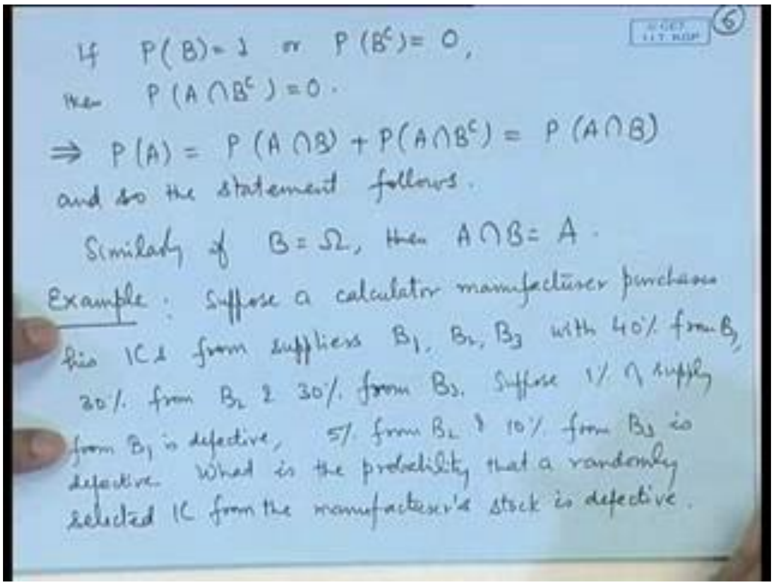
(Refer Slide Time: 15:58)



So, let B 1, B 2 etcetera be pair wise disjoint; that means mutually exclusive events with B is equal to say union B i, i is equal to 1 to infinity. Here it could be a finite union also, it will not make any difference to the actual statement of the theorem then for any event A, probability of A intersection B can be decomposed as probability of A given B j into probability of B j, j is equal to 1 to infinity. Further if probability of B is equal to 1or B is equal to omega, then probability of A is equal to sigma probability of A given B j into probability of B j. So, you can see here, if we treat the event A as A consequence of either of B 1 B 2 etcetera; then the eventual probability that A occurs can be represented in terms of that it has been caused by B 1, it has been caused by B 2 etcetera and the corresponding probability of those causes themselves. We will look at some examples here and but first you let me give the proof of this.

So, consider A intersection B; now A intersection B can be written as A intersection and the representation for B is union of B j, j is equal to 1 to infinity. By applying the distributive property of the unions and intersections, we can write as union of A intersection B j, j is equal to 1 to infinity. Now we notice here that B j's were pair wise disjoint sets therefore, A intersection A intersection B j's will also be pair wise disjoint and therefore, the countable additivity axiom of the probability applies here and you can write probability of A intersection B as summation of the probabilities of A intersection B j. Now by applying the multiplication rule on this simultaneous probability, this becomes probability of A given B j into probability of B j, which is actually the statement of the theorem of total probability.

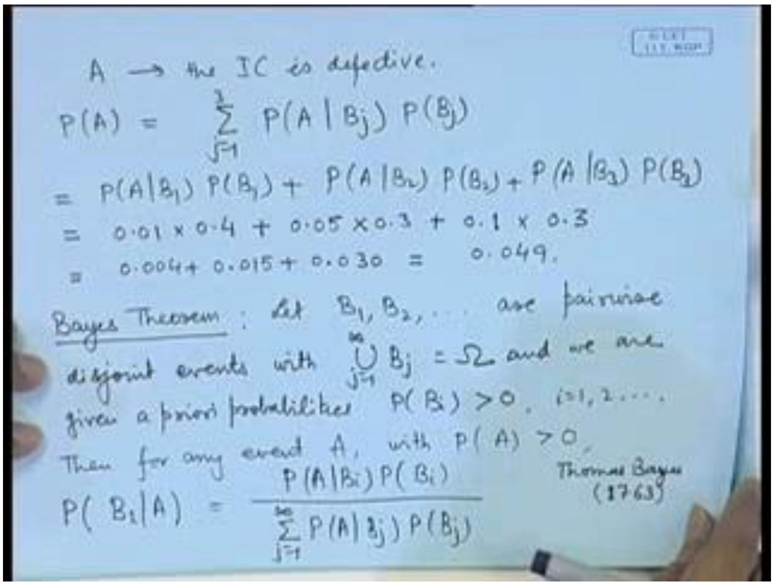
(Refer Slide Time: 19:35)



Now we look at if probability of B is 1 or B is omega case also here. So, if we take probability of B is equal to 1 or probability of B complement is equal to 0 then probability of A intersection B complement is 0 and therefore, probability of A can be written as probability of A intersection B, plus probability of A intersection B complement as probability of A intersection B and so the statement follows.

Similarly, if we are saying that B is omega then A intersection B is simply A and the statement is true. So, let me give an example here, suppose a calculator manufacturer purchases his ICs from suppliers say B 1, B 2, B 3 with 40 percent from B 1, say 30 percent from B 2, and 30 percent from B 3; suppose 1 percent of supply from B 1 is defective, 5 percent from B 2 and 10 percent from B 3 is defective; what is the probability that a randomly selected IC from the manufacturers stock is defective.

(Refer Slide Time: 22:39)



Now, here if we consider the event say A that the IC is defective, then probability of A can be represented as sigma probability of A given B j into probability of B j, j is equal to 1 to 3; that means, since the IC could have come from either first manufacturer or from second manufacturer second supplier or from the third supplier and therefore, the consequence it that it is defective is also the probability is dependent upon who supplied that.

So, this become probability of A given B 1 into probability of B 1, plus probability of A given B 2 into probability of B 2, plus probability of A given B 3 into probability of B 3. So, now probability that it is defective given that it came from the first one is only 0.01, and the probability that is supplied by the first supplier is 0.4; the probability of being defective from the second supplier is 0.05 and the probability that supplier 2 supplied it is 0.3; probability that it was it is defective provided it was supplied from A supplier 3 is 0.1 and the probability of getting the supply from the third supplier is 0.3.

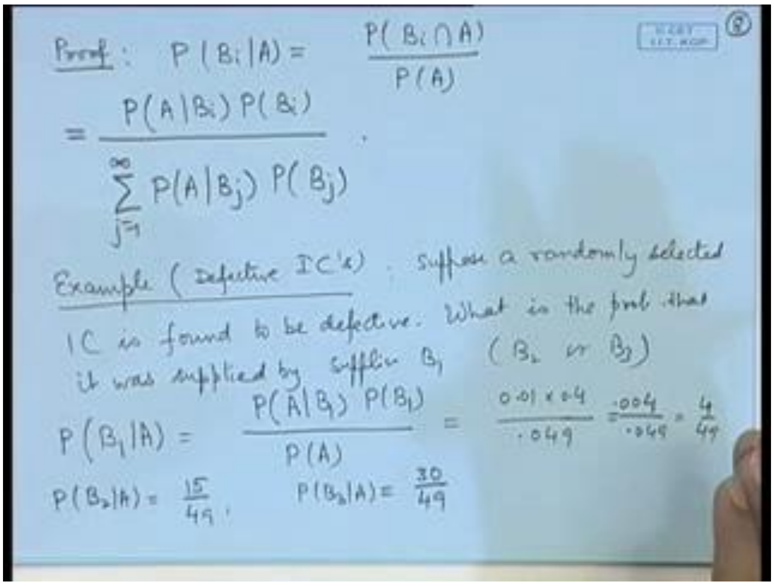
So, eventually this turns out to be 0.004 plus 0.015 plus 0.030, which is equal to so the eventual probability of the IC to be defective is point approximately 0.05; although individually if we see from the first supplier only 1 percent, from second supplier 5 percent and from the third supplier it is 10 percent, but since the procurement is mixed with different percentages getting supplied from different suppliers therefore, the actual probability turns out to be 0.049.

A further you can say consequence of this cause effect relationship is that, sometimes we are knowing the final outcome; now what is the probability that this outcome was caused by something. Now this representation I will consider this way of looking at the probability is called posterior probability, because here you can consider the events B 1 B 2 B 3 that the supplies came from first supplier, second supplier or third supplier as prior events or something which is happening before. Now being defective or non defective is an event which is occurring after that because it is after procurement. Now suppose it says that suppose the IC is supplied and somebody takes it and it is found to be defective; then what is the probability that it came from first supplier or it came from the second supplier or it came from the third supplier, this way of looking at this probability is called posterior events or the probabilities of the posterior events.

So, the result for this is called Bayes theorem, let me state the result first suppose B 1, B 2 etcetera are pair wise disjoint events with union of B j is equal to j omega; that means, they are exhaustive also and we are given a priori probabilities probability of B i as positive, then for any event A of course, since conditional probabilities are involved we must say probability of A, is positive then we define probability of B i given A that is equal to probability of A given B i into probability of B i divided by sigma probability of A given B j into probability of B j, j is equal to 1 to infinity.

So, let us look at the representation of this means that we know the final outcome A, what is the probability that it was caused by the ith cause. So, this is the posterior probability and this base theorem actually helps us to calculate this, it is calculated in terms of the prior probabilities and the consequent probabilities which are called like the cause effect relationship that what is the probability that the event was caused by the ith cause. So, the eventual posterior probability can be represented in terms of this, reverse relationship it is due to Bayes reverend Thomas Bayes, and it was published in his treatise in published in 1763 posthumously; the proof of this statement is let us look at the proof of this.

(Refer Slide Time: 29:33)



So, we will just apply the definition of the conditional probability; probability of B i given A is equal to probability of B i intersection A divided by probability of A by the definition of the conditional probability. Now what we can do is that in the numerator we can apply the multiplication rule in the reverse way, so this becomes probability of A given B i divided into probability of B i. In the denominator you have probability of A applying the theorem of total probability, we can write it as sigma probability of A given B j into probability of B j, j is equal to 1 to infinity, which proves the statement of the Bayes theorem.

one thing which one can notice here is that the proofs of the statements of the theorems are quite simple, this is happening because we have given an axiomatic structure to the to the probability and therefore, all the statements are simply consequences of certain relationships in the set theory, when the axiomatic definition was not known all of these statements had complicated proofs, because if you use the relative frequency definition then you have to look at the various relationships between the limits; if you use the classical definition then you have to represent all the events in terms of favorable number of cases and the total number of cases etcetera. So, all of these statements used to be quite complicated the major contribution of the axiomatic foundation is that it made all the things extremely straight forward and also valid because you can everybody can verify that the statements are true.

Let us look at the previous example of the IC defective problem; suppose a randomly selected IC is found to be defective, what is the probability that it was supplied by say supplier B 1 or B 2 or B 3. Let us look at the first one; in the first one we are interested in finding out the probability of B 1 given A. So, this is by the definition of Bayes theorem, it will become probability of A given B 1 into probability of B 1 divided by probability of A. We already evaluated the probability of A as 0.049. So, it becomes simply probability of A given B 1 that is 0.01 into 0.04 that is the probability of B 1 divided by 0.049, which is equal to four by as it 0.004 by 0.049, which you can write as 4 by 49. If you look at probability of say B 2 given A then in the same logic this will become equal to 15 by 49, if we calculate probability of say B 3 given A then by the same logic it will become 30 by 49, you can see here the difference in the posterior probabilities.

Although only 30 percent of the product comes from manufacturer third, but if the finally, selected IC is found to be defective then it is most likely to have come from the third supplier. The reason being that in the hypothesis of the problem, it has been assumed that more number of products from the third supplier are defective, which is actually 10 percent whereas, from the first supplier it is only 1 percent. So, although more supply comes from the first manufacturer; however, the probability that it came from him provided it is defective then the probability is much more small.

So, this is an important you can say consequence of the Bayes theorem, B 2 given is A 15 by 49, probability of B 3 is 0.3 whereas, probability of B 3 given A becomes 30 by 49 which is much higher. So, the probabilities get revised in the light of the knowledge about the outcome.