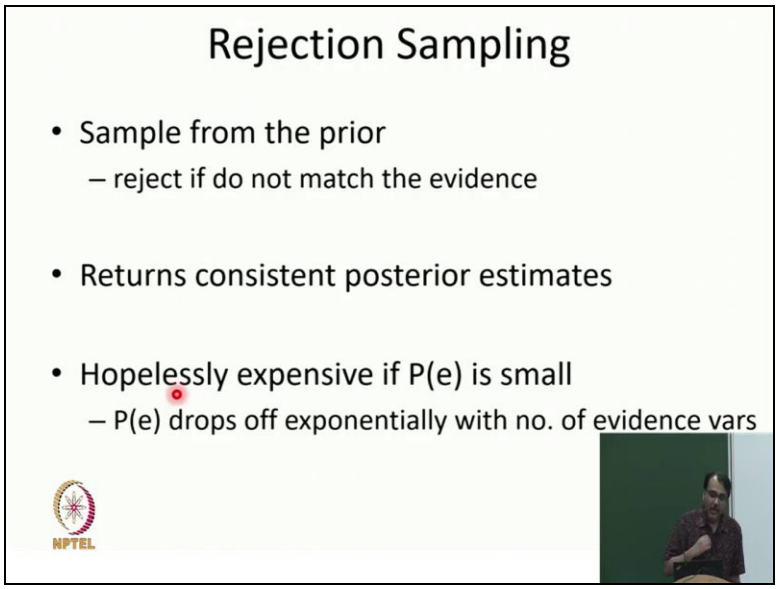
Artificial Intelligence
Prof. Mausam
Department of Computer Science and Engineering
Indian Institute of Technology - Delhi

Lecture-59
Bayesian Networks: Likelihood Weighting

What is the drawback of rejection sampling? Can anybody come up with the drawback, Purva, yes, very good Purva says newscast is actually is element for the purpose of B given C. So, let not even sample irrelevant it is a good point you can get rid of the irrelevant variables, fine. What else? There is one more sort of more important in some ways, yes, what is your name? Abyutha yes, if probability of call is very low very good. Let us think about what is going to happen.

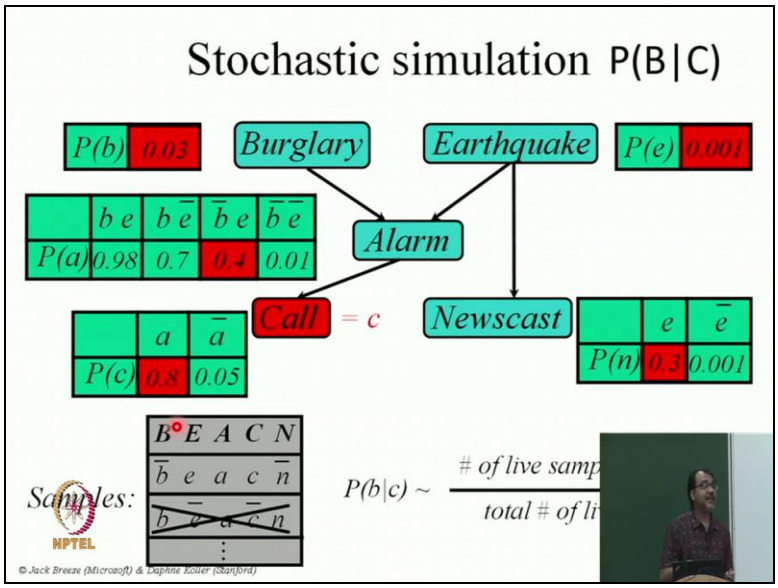
Probability of call is very low, now we are interested in B given C but C itself is very low. So, what is going to happen? Most samples are going to get rejected. It is going to become an extremely expensive algorithm.

(Refer Slide Time: 01:22)



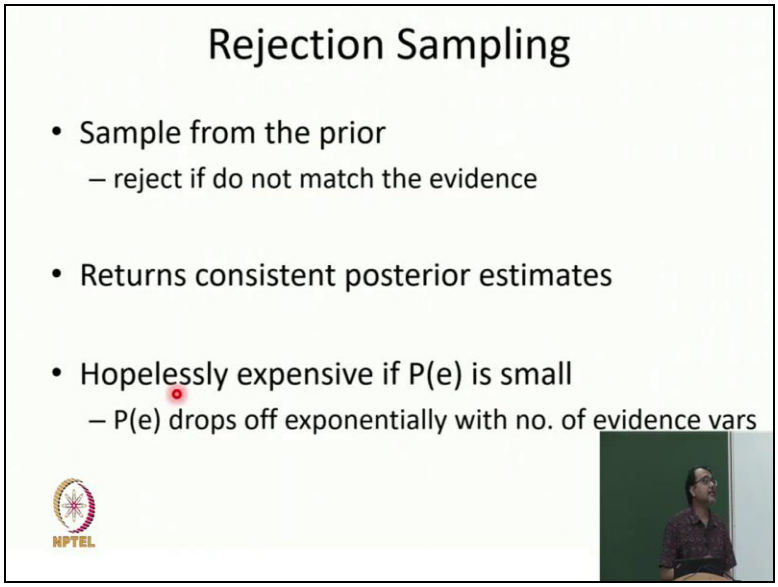
So, this is what is rejection sampling it rejects if it does not match the evidence it returns consistent posterior estimate but if probability of e is very small then it is going to be hopelessly expensive and moreover and this is important point probability of e drops of exponentially with number of evidences can drop of exponentially with number of evidence variables.

(Refer Slide Time: 01:59)



So, if I have too many evidence variables, like my question is probability of query given e1, e2, e3 upto en and all there so many evidence variable and I am saying e1 should be true and e1 should be false e2 and should be true whatever. So, I have given a very specific assignment of n or k evidence variables. Then what is going to happen is the probability that I am going to get is exactly those evidence variable in that specific configuration is going to be extremely, extremely small.

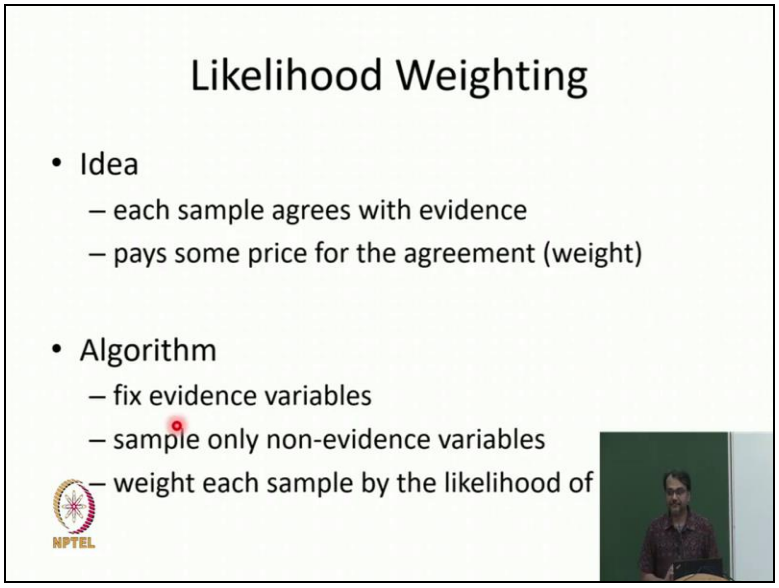
(Refer Slide Time: 02:23)



Most of the Sample are going get rejected. So, that is the issue with rejection sampling. So, intuitively what do we want in our sampling procedure? Intuitively what kind of sampling

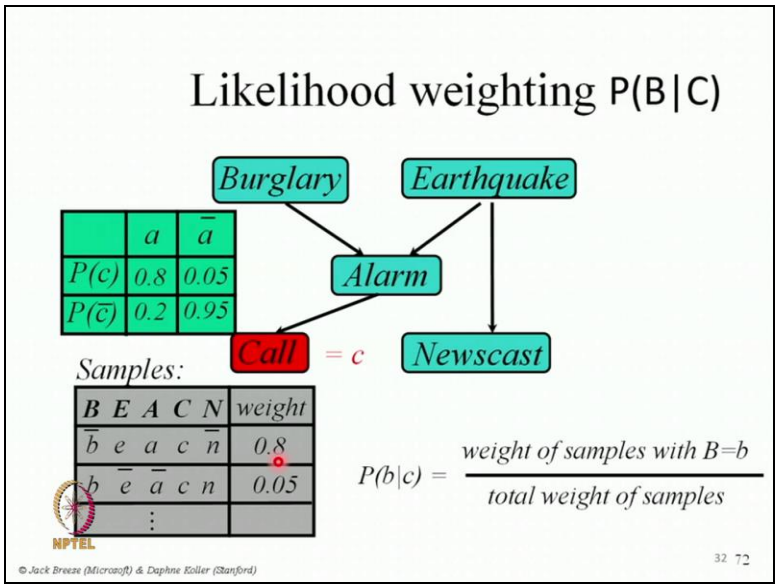
procedure would be better sampling procedures, which, in which evidence is always true. Very good. Veyanses says look, I should not be rejecting anything. My evidence should always be true otherwise I am wasting time. Show my goal is now to come up with the sampling procedure evidence is always true ok and one such algorithm is called the likelihood waiting algorithm.

(Refer Slide Time: 03:05)



Now the problem is we have to force the sample to agree with the evidence. And if we force the sample to agree with the evidence we will not be sampling from the power distribution and we will not be sampling from the unbiased distribution, so we will have to add correction to our algorithm. Ok and this correction would happen in the form of the weight. Ok let us look at the specific example.

(Refer Slide Time: 03:34)



So, we started with; we started sampling burglary was false earthquake was true, alarm was true. Now come is the key step are we going to sample again from call given alarm. What do we want? We want C to be true in this sample. We do not want to reject anything we want C to be true. Tell me in what fraction of the cases will it be true and in what fraction the cases will it be false? In point 0.8 fractions of the cases it is going to be true.

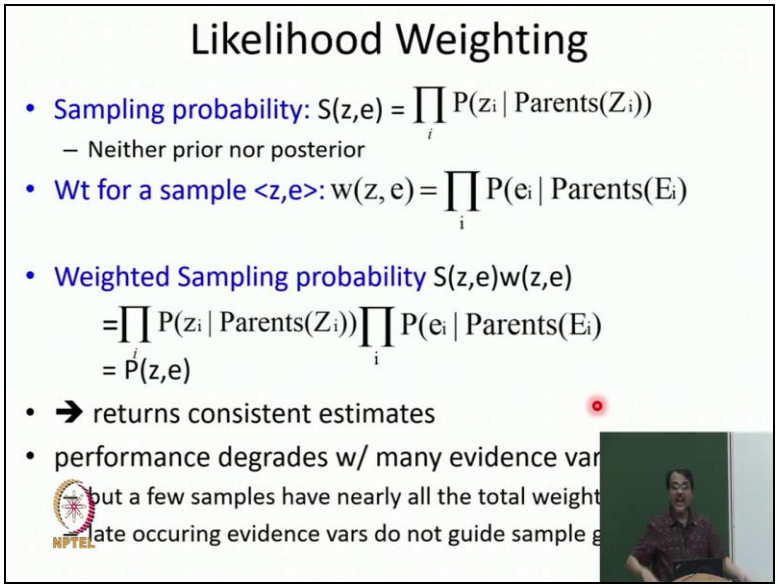
And in 0.2 fraction of the cases it is going to be false. If we had rejection sampling you would have sample this and 20% the sample would have got rejected. Actually, I do not sample it. Then this sample is only worth 80% it is not worth the full 100% because we have already rejected some. So, we are going to say that C is true we are not going to sample it but with the weight of 0.8. This sample is only 0.8 of a sample. If this sample could have been one if we had allowed to sample, you know C and C bar but we did not.

This sample is only wroth pointed. If there was another evidence variable later then that weight is going to get multiplied in the first way. We are going to keep multiplying the weights and then eventually we are going to say probability B given C is the total weight of samples, where this true divided by total weight of all the samples. This is the weighted average not each sample is equally important.

Some samples really aligned with the unbiased sampling like 0.8 but some sample did not we have to pay lot of cost because here C given A bar would have been estimated with; would been sampled with probability 0.05. So, 95% of the time we were not has seen this sample only 5% of the time you would have seen this. So, its weight is only 0.05. And this algorithm is called the likelihood waiting. Now only in 5 minutes without going into too much detail, I will show you this actually gives you the right probability distribution.

And if you do not get this is ok you read the book, but I will give intuition of why this is a right algorithm, you think about that. See there are two things that are going on. One is that I am doing the real sampling but one is that I am not doing any real sampling but adding some weight. Together they should be equal to the sampling from the prior distribution. If together they are equal to the sampling from prior distribution we are good right.

(Refer Slide Time: 06:43)



Now, what is the probability with which we sampling the probability sampling is only probability of hidden variable is given parents of hidden variable, enquiry variable so non evidence variables. So our sampling probability, right only samples real tosses the real coins for only those variables which are non evidence. This is by the way neither prior nor the posterior. In prior they would have added probability of evidences also.

And in posterior they would have been sampling each variable given the evidence variable. We are not doing that. We are always sampling a variable given its parents. It is neither prior nor the posterior somewhere in the middle. But what is the weight of a sample? Weight of the sample is probability of the evidence variables given its parents. We sampled it specifically but for that evidence we use the weight that was in the conditional probability table and we multiplied.

So that is it now since you are doing weighted average let us look at what the product of these 2 terms. Product of these two terms is nothing but the prior probability distribution. So, we thing about it we have divided our problem into two parts one when we are doing real sampling and one when we are doing sort of implicit sampling and rejection, but we are putting that is the weight in the term so that we do not actually have to do it and reject.

Together the sampling probability and the weight is nothing but something from the unbiased prior probability distribution, therefore, it returns consistent estimates. Now come the more interesting question for you guys, think intuitively again. Is likelihood writing a better algorithm than rejection sampling? Better in what sense, good questions better in which sense is it version in some sense? We need less number of samples for; we would need less number of samples for because we are not ejecting anything.

We are still getting consistent estimates. So, we are not hurting on the quality or the accuracy of the algorithm, but we are reducing the time taken. So, overall is it a better algorithm? Yes, better in that sense but we are still not satisfied. What is the problem with the likelihood weighting this is not easy, but let us see if somebody has any intuition about what is the problem with the likelihood weighting algorithm.

It is only sampling evidence directly. It is never sampling something that needs to be rejected but still it is not the greatest algorithm. Why? What is the fundamental issue with this kind of algorithm? What is your name? Videsh says deciding weights is that a problem in this algorithm? Weight is whatever the probabilities of the conditional probability table of the evidence variables. So there is no issue on deciding let us make sure that you understand this.

This weight is pointed because this number is pointed. This rate is 0.5 because this number is 0.5 and that depends on whether I sample alarm to be true alarm to be false. So now deciding weights is not a problem. Can somebody else see what might be an issue with the likelihood weighting algorithm? Sukriti yes, ok Sukriti says weight are multiplied and overtime some samples will have very small weight.

Ok how would that be a problem? Very good so Sukriti's hypothesizes that it is possible that there are some samples with a very high rate and every other sample as an extremely small weight and in fact, that is sort of what is observed in the likelihood weighting algorithm as well. Let us take another step back when we want to compute probability of B given C what probability distribution do we really want to sample from?

What probability distribution do we really want to sample from? The probability distribution in which C is always true we are only interested think about this suppose I have got a call, better to think intuitively here. Suppose I have got a call from my neighbour that there was an alarm. Probability of burglary should not be again sample with probability 0.03 and probability of earthquakes should not be; again sample with probability 0.001 that probability distribution has changed.

If I got a call there is a good chance there was alarm. If there was alarm there is a good chance that either burglary or earthquake, but when I am sampling I start sampling with prior distribution. Initially I am not sampling doing any change in my distribution when I am sample I continue sample in the prior distribution until I get evidence. Now the subsequent variables I do sample evidence in mind. But I have already done the wrong thing. I have already sample lots of variables and come into a part where it is not even clear whether the evidence would have happened in a high probability or not.

Really what should I have done? I should have sample from the posterior distribution. But I am sampling from somewhere initial in the prior distribution and later ends in a semi posterior distribution modes and that does not necessarily take me to the regions which are interesting and important for the given evidence. So, Sikrutis point is right that there are very few samples that

nearly have all the total weight all the samples low weight and therefore likelihood weighting algorithm does not do very well.

Moreover most importantly to me late occurring evidence variables do not guide sample generation and those are what should be guiding sample generation. If I start out by saying let us see, whether there will be burglary. Let us see whether there will be earth quack the false and false almost probability 0.97 this would be false probability 0.999 this would be false and very, very high probability both of them will be false.

So, 98% of the samples would be on 96% of the samples would be in the region where neither burglary happened nor earthquake happened and I am always thinking it is low probability I would have got the call very low weight of the sample very low weight of the sample and only by very small chance then I will have a higher weight of the samples but that does not explore the entire Bayesian network very well. I really as soon as have call I should increase the probability of sampling burglary and earthquake and that was not happened.

And so now finally we are going to study the algorithm that has the characteristic where when each variable is sampled it is always mindful of the evidence that is given to us. Skruthi asks where does it converge; we just proved it quickly the weighted sampling probability is the prior probability distribution and we can answer any query with prior probability distributions. So, in the limit infinite samples does it converge?

Yes, in the limit of infinite samples it will converge then you think about it. In practice the weight are off probability are off it takes long, long time to converge because the number of samples need is very large if evidence variables are at the lower part of the tree not at the other part.