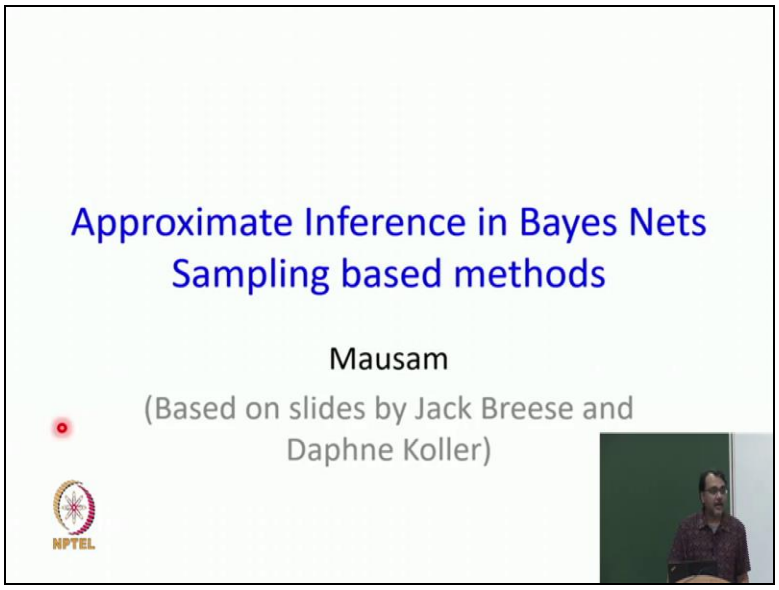
Artificial Intelligence
Prof. Mausam
Department of Computer Science and Engineering
Indian Institute of Technology - Delhi

Lecture-58
Bayesian Networks: Rejection Sampling Part-6

Alright, let us get started. Up until now we have looked that the representation of Bayesian network, the syntax, the semantics and one inference procedure and exact inference procedure. The reason I call it exact is because it will give you the exact probability of a particular query and exact inference procedure that solved answers to the queries at the tip of the Bayesian network.

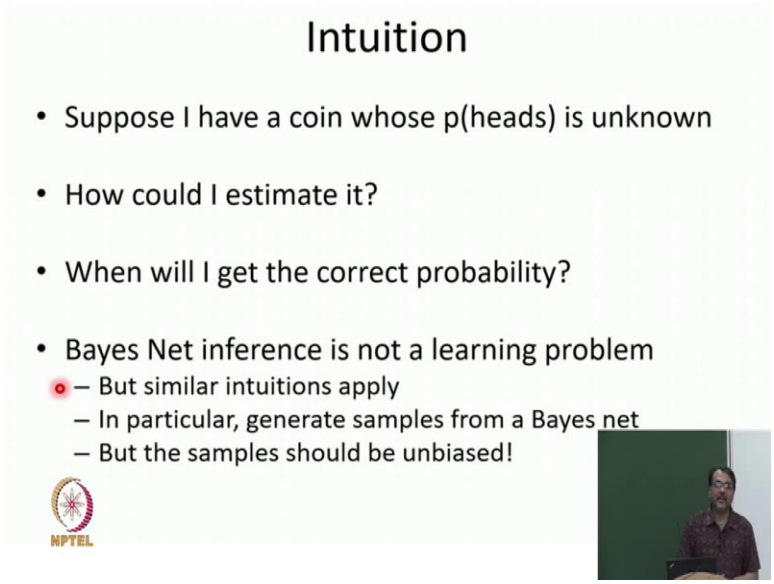
Now, of course we have also studied that Bayesian network inference is partially complete which basically means that we cannot hope that will be solved in a practical time for large Bayesian networks. And so it is important that we come up with some kind of an approximate procedure and we are going to study one approximate procedure, one that I am very fond of, again it is a sampling based method. We are going to use sampling to answer queries in a Bayesian network.

(Refer Slide Time: 01:27)



And we will get started with a very simple intuition; intuition sort of that is not exactly appropriate for the task at hand but will give you would starting point.

(Refer Slide Time: 01:35)



Suppose I want to estimate probability of heads for the point, whose probability is unknown. We can hope that maybe it is biased we do not know that. If you are in such a situation then we have to estimate probability of heads, we do not know the probability but we are given the coin. What are we going to do? Tossing many times, take lots of samples from the coin. And then say, out of the total number of tosses, what fraction of times did I get heads and that would be my estimated probability of heads right.

To estimate any probability, we can sample we can take some person take help from samples and we can use some kind of a fraction to compute that particular probability at hand, ok. This is a very general concept to estimate any kind of probability we can use sample. And of course, when will I get the correct probability? Obvious question, when will I get the correct probability for this? When I sample lots and lots and lots of times in the limit, the number of samples sent to infinity, right.

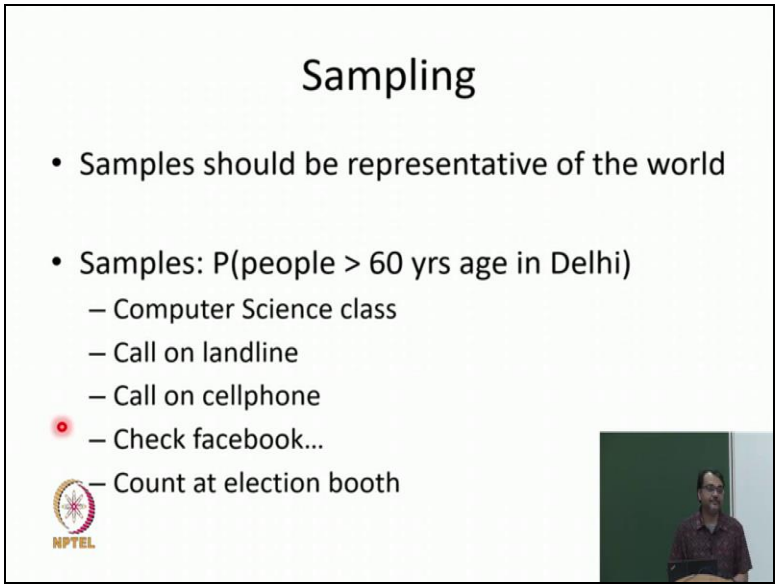
We now know that we have a simple approximate procedure to sample probability of heads. The first probability of heads is 0.5. Am I going to estimate 0.5 in a set, in a 10 samples, maybe not, unlikely? You might not have 6, 4 splits even 7 to 8 split by Chance, you might also have a 5,5 split, but you have to be very lucky to happen. So, this would be an approximate computation but after is approximate computation as we keep increasing our number of samples the probability

will 10th closer and closer to 0.5. You might have a million sample for head million hundred sample for tails, ok.

Now, Base net inference what we have been doing. This is, this is really a learning problem. We are learning the probability of X by trying it out. Technically Bayes net inference is not a learning problem. But the same intuitions apply in particular like we can sample a coin, believe it or not, it is very easy to sample from a Bayes net and the samples will be unbiased or can be unbiased if you try right?

Now, what is the meaning of bias in samples. Let us first make sure that we understand that. So, bias or unbiased samples are the samples which are the representative of the world or in this case the representative of the Bayesian network. For example, suppose I want to estimate the total probability of people greater than 60 years age living in Delhi.

(Refer Slide Time: 04:37)



This is my goal. We want to estimate the probability of people greater than 60 years age, living in Delhi. What factors the people who live in Delhi are there in 60 years age? Now, suppose I come to the computer science class and I just check how many of you are us are greater than 60 years age. And if you study probability is zero even I am not 60 here. Even though I may look like one.

Then I did use that of because in the computer science class there is nobody is there greater than 60 years of age so the probability is zero. Would this be a biased sample or an unbiased sample? May be a biased sample, I could call on a landline would this mean less biased? Probably less biased but would it be biased also? Why did you say it is going to be biased? If you are really calling random people then what? It would not be bias. It would not be biased. Why?

Somebody can tell me why it would be Biased yes, Nishant? Children under 5 may not be picking up the phones, very good and another reason it might be biased using landline the higher chances are the old because the youngsters do not use landlines anymore. How many have a landline at home? Very few, may be less than 50%, if you had this question 10 years ago, everybody would have raised the hand, almost everybody.

Then there is a rich food bias right, a lot of people who may not have any phone or at least a landline or the people who do not have a study home or do not have a home which is connected to a landline phone. You can call on a cell phone. That might have another kind of bias, people must have started living in Delhi but they moved to Bangalore, right or things like that. You can check the Facebook pages. That is probably less biased. We know all the old people are now on Facebook which is why none of you are. Where are you guys these days?

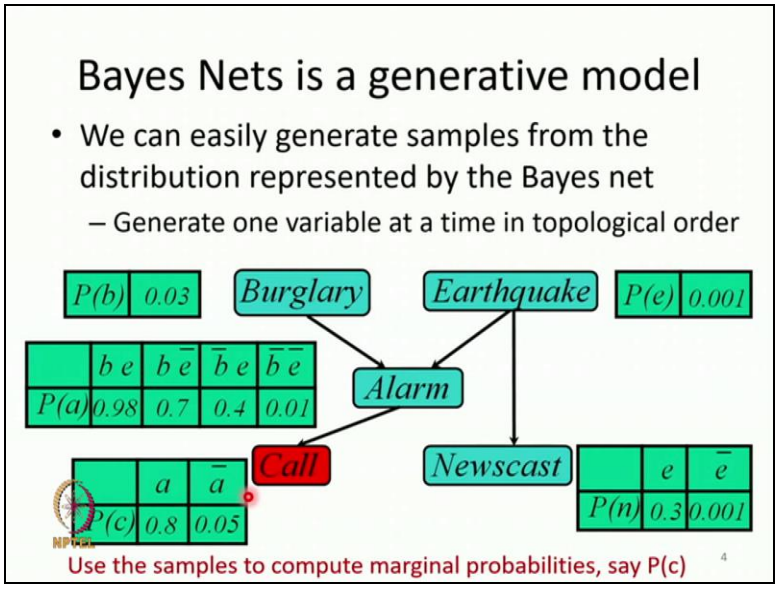
Instagram that would have been told. If you are still on Facebook and not an Instagram think about your age, Ok. Or you can do a count on election booth whatever kind of sample you choose. Whatever kind of sample you choose, it is going to be that is my all these people remember when Arvind Kejriwal government was not formed and people were doing this exit polls. I do not know if you remember it is 5 years ago.

Almost half the people predicted that BJP is going to have half the service predicted and A, B going to win but the margins and was very low. Kejriwal was getting 7 out of 76. Why because they were doing some kind of a Biased Sampling and in that particular case, sampling became too bias than people who came for voting and the people who responded exit polls became today. And because we are in the Bayesian network world, the good news is that we can actually deal with bias explicit.

So this is just a general idea of what does it mean to have a bias sample and an unbiased sample, but we are given the model of the world. The model of the world is the Bayesian network. Because you are given the model of the world, we will adjust focus our energies on sampling from the network itself and not to worry about many of these Real world problem. Of course, the model of the world may be wrong. The Bayesian network itself is very wrong.

But that is ok we have approximated things out, we have abstracted things out, all models are wrong, but we believe that the Bayesian network given to us is useful. We are going to work with the network that we are given with the conditional probability that are given to us. However, they may have been estimated. They may have been errors in the estimation we are not going to worry about that particular bit. Ok.

(Refer Slide Time: 08:50)



Good news is I am going to mention or introduce a new term not in detail. If you take my healthy class we will spend 3 or 4 lectures, we will just be talking about these two types of models: one is a generative model and one is a discriminative model. Surprised, somebody knows it in the class. Discriminate model cannot generate samples from the distribution, but a generative model can generate samples on the distribution network. Ok. Bayesian network is a generative model.

You can easily insert extremely easily general samples on the distribution represented by the Bayesian Network and network represents joint distribution, right you always call this and what is the algorithm. The algorithm is so simple, generates one variable data at a time into logical order. This is a good example, they are going to use this one example, Burglary, earthquake and alarm and let us say John and Mary not separated anymore and then there is one neighbour also does not call.

And then there is something on the radio or the newscast which tells you whether there was earthquake or not. We are using this particular example and we are given this conditional probability tables and what we are going to do is we are going to just sample from the distribution, what does that mean? We will toss a Coin with probability 0.03, we will say burglary is true and probability 0.97, burglary is false.

By the way, how do you toss a Coin such that it gives you two things with probability 0.03 and 0.97, again this should be obvious people who actually do some coding in life, but for others, how do you toss a Coin in any programming language, what function do you use, User rand function. So, rand function always gives your number between 0 and some rand max minus value, Right. And so we can take this dividing line and then we can say that ok our total length is rand Max 0.03 rand max is one side of the coin and 0.9729 rand max is the other side of the coin and we have go in our threshold.

So, sample a random number between 0 and -1 we assume that it is a uniform sample and then we say whether it is less than the threshold or greater than threshold located in the such a give point. So, random number generator is always gives us a uniform sampling. We are all the pseudo random. There are expected to give a uniform Sampling and then we use the uniform sampling with the right thresholding to give us the probability 1.03, right.

So, we crossed a point where the value of 0.03 so that we can figure out the value of burglary and earthquake. We cross a point where the value is 0.001 we figure out the value of earthquake. Now, we have the value of burglary and earthquake. The alarm probability is determined by one of these four values depending upon whether burglary and earthquake but true burglary was true

earthquake is false and then let us say burglary was false, earthquake was true, then we toss a Coin with probability 0.4 and that tells us whether actually alarm goes off or not.

Based on the probability of alarm we figure out whether somebody called us or not. So, in other words, we have done this small probability flip and at the end of all the flips, What do we have? We have one sample from the joint distribution. He is false, he is true is true. E is false, is false whatever, that gives us one 7 samples, we can do this coin flip many, many times and we have got lots of samples and notice, each sample is generated by the joint distribution.

So, these are real samples, real mean these are simulated samples, but from the distribution of the Bayesian network represents. And therefore we can use this samples to computer any kind of a marginal probability. Let us say want to compute probability that somebody going to call us without anything any evidence. So, we can just check in how many samples was see through and what was the total number of samples and that gives us the probability that somebody gonna call. I am going to pass this. Is this ok for now? It is like the same thing.

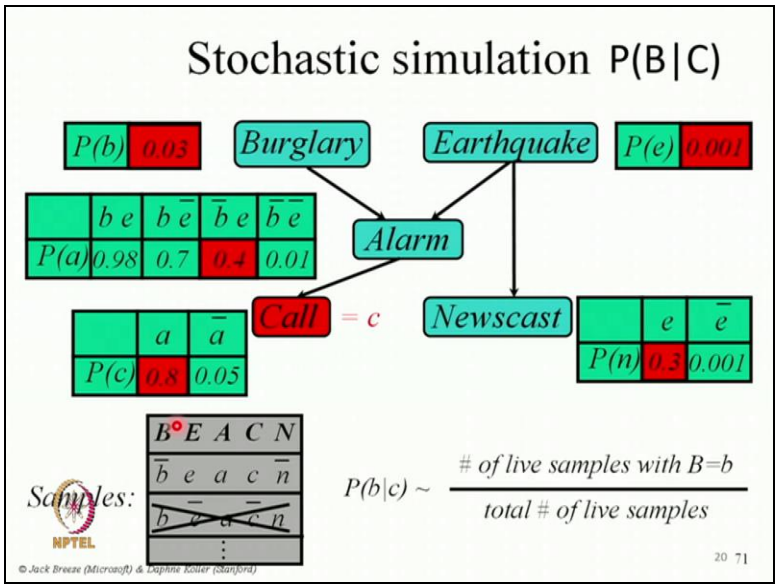
How do we compute probability with sample of toss the coin many times how many times we get head. In the same way, a Bayesian network represents the joint probability distribution so we sample from the joint. What is the sample, joint sample is each value filled up. Now, we want to estimate the marginal probability of somebody going to call. Then in some samples call is true in many examples call is false, the total number of samples in the denominator and number of samples which calls to the numerator and that gives us the marginal probability that somebody calls, ok. Any questions?

If you did not understand this then the whole next part of the lecture is going to be gauzy for you. So; any questions? How do we generate a sample? That is a question how do we generate a sample, we go in the topological order. We start with the node which has no parents. No its value can be determined by independent coin flip rights example, Whether Burglar is going to happen or not also we can toss the coin 0.03 and burglary at 0.97 is available.

Similarly, earthquake we do the same thing. Now in this sample burglary has been determined an earthquake has been determined that is what happened so far. Will alarm go off? That is determined by the specific condition probability. Based on what values we have sampled for burglary and earthquake. Let us say sample that burglary happens but earthquake does not happen. So then we will toss a Coin and with 7 probability so that the alarm will go off.

So, we are estimating the various possible worlds that can happen with the Bayesian network but not only are they estimating them, we are estimating them such that the world that is more likely will happen with a higher probability. That is important.

(Refer Slide Time: 15:29)



Now the more interesting question of USB is how do we compute? Let us say we get a call the probability that burglary has happened. How do we estimate probability of burglary has happened? We have b given c? Any suggestions Let us do the again the 3rd experiment, we are standing and cannot place that you are interested in. What is the probability that a person has a Masters degree given that their age is better than 60 years, ok?

What is the b given c, what is the probability that they have a degree? Graduate, postgraduate degree or more given that their ages get in 50 years. What do you do? Standing in Connaught Place, what do you do? You go to a random person and first ask, what is your age? Is your age

greater than 60 years? What is your name, Shahn. Shahn says is your age greater than 60 years that is the first question I will ask.

Now, if the ages not greater than 60 years, what do you do? You just ignore this you are not useful to me, please go away. We reject the sample. If you are interested in probability of b given c and c is false. That sample is not useful to us. We reject the sample. However, if their age is greater than 60 years then what do we ask? You have a Masters degree and if they have a Master's degree, we increment the numerator by 1 and either we have the values incremented in your denominator and that is the probability.

Now, if you want to compute probability of burglary given that we are going to get a call, what do we, what we do? Given that we got a call. Now you have intuition. What do we do? We take, let us sample from the joint distribution as you are doing in the last slide. And now what do we do? There are some samples when call is true and there some samples were call is false. What do we do? We reject all the samples the call is falls. Guess what is this algorithm called?

It is called Rejection sampling. Engineers do not have very creative mind when it comes to naming an algorithm. By the way there is a trick in naming algorithm. You have to make sure that the name is funny or at least catchy and you have to make sure that it does not exist in Google with something else. Take one of my students have named the latest data set as KA. That is interesting as a name but as soon as somebody searches Ka, they are going to go everywhere.

The amount of Ka deposits you know and global warming and all that just, terrible name for a system. So when you ever get to a point where you are to name a software or in some cases name a system that is a research prototype always work very hard on naming it. You have work thousand hours in coming up with the algorithm. Please spend one to two hours in naming it, alright. So that is exactly what you do.

You start sampling you get one sample like you for sample B, B was false in example, e, he was true, e was true unit sample, now B false, e true so you are going to use the probability for let us

say an example a let us say is true. Then, your next Sample C with probability 0.8 because a has been through, you get C then unit sample and with a probability 0. 3 you get n false.

Now, is the sample relevant to you for probability of B given C? Yeah, this is relevant. Similarly, if we get this other sample is this relevant to us? No, we do not care for this. And so we reject. And so then we want to finally compute probability B given C that can be computed as number of live samples with the burglary was true, given the total number of live samples. We have already rejected. This is called Rejection sampling.