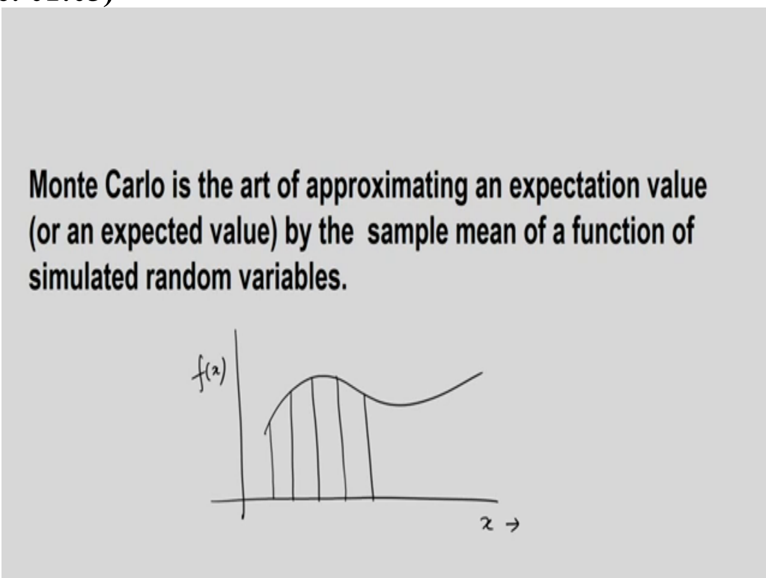
Numerical Methods and Simulation
Techniques For Scientists And Engineers
Saurabh Basu
Department of Physics
Indian Institute of Technology- Guwahati

Lecture 20
Monte Carlo Technique Importance of Sampling

So let us briefly review the Monte Carlo technique that we are learning here. And we will introduce what is called as an important sampling. This is a very important you know development in the distribution of the random numbers which would eventually give a very accurate a much rather much more accurate estimate of the integral that we are attempting to compute.

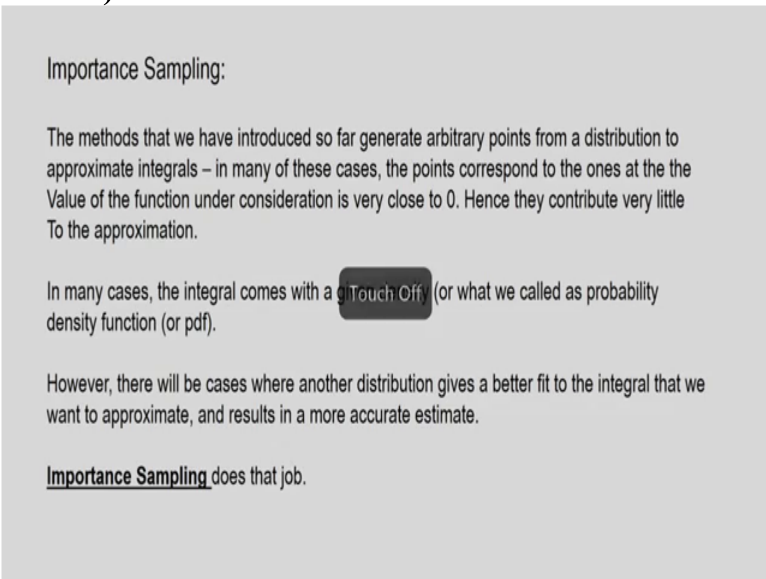
(Refer Slide Time: 01:03)



So in summary Monte Carlo is the art of approximating an expectation value or we also can call it an expected value as we were calling during our discussion by the sample mean of a function of simulated random variables. So instead of choosing grids of definite width we actually choose random numbers. So say we need to integrate a function like this and so this is f of x as a function of x, so instead of doing such grading which we have d1 for these known integration techniques here we sort of and so on.

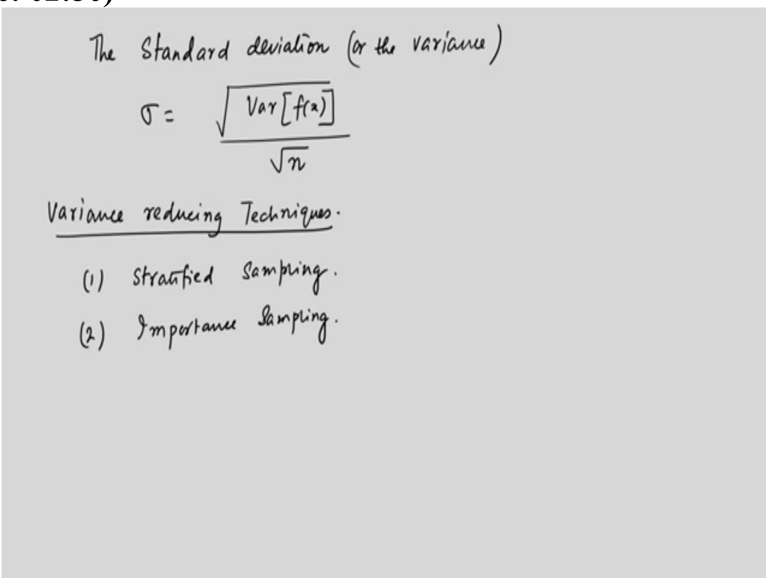
So here we choose random points on this f of x which is the curve here and then calculate it from its sample mean so these random points are taken from some distributions it could be uniform distribution or it could be a Gaussian distribution or a normal distribution. And we actually calculate the value of the functions at those points and take a mean of that and that is the expected value or rather the value of the integral. So this is the main idea behind the Monte Carlo.

(Refer Slide Time: 02:31)



So let us look at what random sampling is or rather important sampling is but before that let us do another one.

(Refer Slide Time: 02:50)



We will come back to this in a while. But let us remind you that the standard deviation or the variance of the MC technique is given by this Sigma equal to root over variance of these f of x and divided by the root over n where n is the number of sampling points that we have drawn from a random distribution . And we know the, you know the merits of the Monte Carlo technique. It is free from the curse of dimensionality so that if you go from 1 dimension to dimension equal to 1 to dimension equal to 2or dimension equal to 3 and so on.

It will, does not hurt the usage of it or applications of it and of course as we have said earlier that our sole idea is to reduce the variance. And if we reduce the variance and then we get more and more accurate values. And for that two things can be done. Either we increase the numerator in which Sigma gets decreased; so increasing the numerator means that increase the number of sample points.

Now this is something that the experience says that the convergence is very slow. So if you increase you know n from 10 to the power 6 to 10 to the power 8 or 10 to the power 9 that is the number of sampling points that you have chosen from a random distribution then the convergence or the accuracy, improvement of accuracy is not all that much. Whereas if we can reduce the variance that is decreases the denominator, then, Sigma will decrease as well.

And this seemed to be a more effective way of tackling the problem and this is what we want to understand and these techniques for which we increase the accuracy is called as a variation of variance reduction or reducing techniques okay. And the technique this particular has a number of you know techniques there. And we will talk about mainly two of them. One being what is called as the stratified sampling and importantly what we will learn is that what is called as the importance sampling.

And in fact the important sampling is the more important of these two and we will discuss it with a little more elaborately. But let us also look at the stratified sampling and understand that how it works. We will come back to this the important sampling.

(Refer Slide Time: 06:07)

Stratified Sampling

This technique divides the full integration range into various subspaces.

The final result is the sum of all partial results.

We divide the integration domain, R into K regions, namely R_i; $i = 1, 2 \ldots K$. The expected value is,

$$E[f(x)] = \int_R f(x)\,p(x)\,dx = \sum_{j=1}^{K} \int_{R_j} f(x)\,p(x)\,dx$$

The MC estimate of the expectation becomes

$$I = \sum_{j=1}^{K} \frac{Vol(R_j)}{n_j} \sum_{i=1}^{n_j} f(x_i).$$

n_j : number of MC points used on R_j

$Vol(R_j)$ is the volume of the subspace R_j.

So what is this? So this technique divides the full integration range into the various subspaces. So the final result is the sum of all partial results okay. So how this works is that so we divide the integration domain let us call it R into say K regions. So namely, let us call them as R I equal to so this is the these are RI where I runs from 1 to K. So then, the expected value in literature this is often called as the expectation value or just the expectation okay.

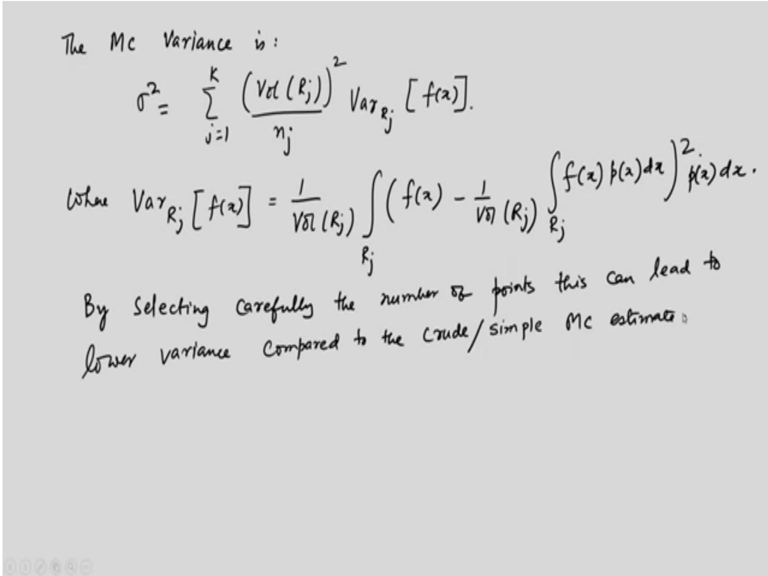
This means the same thing the expected value is written as this E and then f of x which is equal to over this space Fx with a probability density function P of x and now since we have broken it down into K regions for evaluating the integral we will have to sum over these K regions. It is not n sorry about that so this is K and then we integrate over a single you know jth block or jth subspace. And then you sum over all the subspaces okay. So, this is so, the the Monte Carlo estimate.

We call it just MC so MC estimate of the expectation becomes I equal to j equal to 1 to k volume of the jth subspace this nj is the number of sampling points in the jth subspace and then I equal to 1 to j f of x i okay. So that is the Monte Carlo estimate of this integral which in addition to the f of x i from I equal to so this is n j rather. So from I equal to 1 to sum n you also need to have another summation.

The over this all these subspaces where is the number of subspaces go from 1 to k so the entire region R is actually broken into K subspaces. So just to remind you that n j is the number of MC points used for the integration purpose on R j. So on this R j and the volume R j is the volume of the subspace R j much. So this looks simple because it just gets extended from 1 you know region or I mean the entire region being considered with the same probability density function from that we go to a K such regions, which the entire region of integration is broken into K of them.

And then doing the procedure, repeating the procedure for all of them and then taking the the sum of all these subspaces. So if we claim that this is any advantage over the crude or the simple MC estimate that we have done, then, we have to convince ourselves from calculating the variance.

(Refer Slide Time: 11:34)

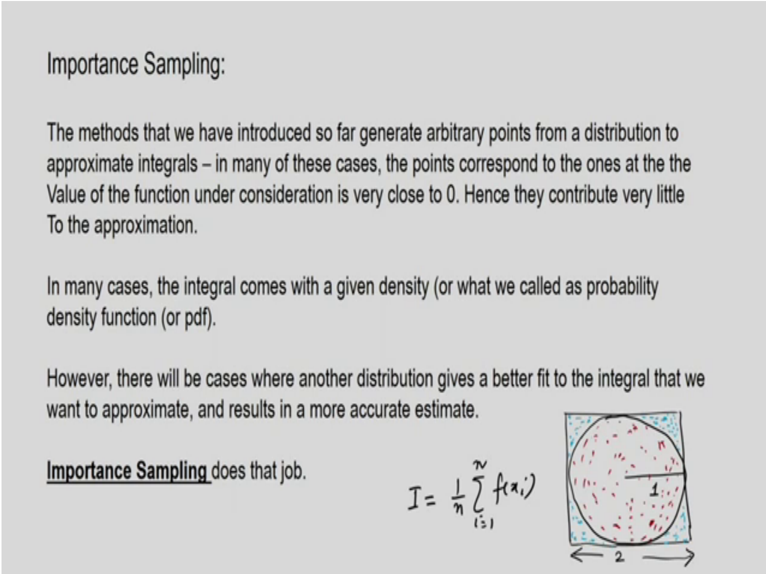


And so the MC variance is so it is a sigma square equal to sum over j equal to 1 to K and a volume of Rj and so this is a square divided by n j. And then the variance of this R jth subspace and this f of x so were these variants of this R jth subspace for the function the integrand. In fact it is equal to 1 by volume of the R j. And then of course the usual variance, that we talked about 1 minus volume of R j. And then, you have a Fx px dx whole square.

So this is integral over RJ and Square and then of course we have a px dx okay slightly longest expression. But it just says that you know this is how the MC variance is calculated in this stratified sampling technique and so if you select carefully the number of points then this can lead to a lower variance than the simple-minded MC that we have seen earlier. So we will write that by selecting carefully the number of points this can lead to lower variance compared to the we have used this word earlier crude or a simple MC estimate alright.

So the next technique that we are going to consider is called as the importance sampling let us go back to our slide which is there.

(Refer Slide Time: 14:23)



So this important sampling is really believing that not always you have a unique probability density function would give you an accurate estimate. So in some cases you could have a better probability density function which gives a better estimate they could be proportional to each other or they could be you know different than each other. But we will have to make sure that that gives you a sort of better estimate of the integral.

So if I am trying to sort of understand what is this importance weigh? let us see this old problem that we have done. So we have a circle consider this as circle, something completely inscribed it is hard for me to draw this. But let me see if I can. So we wanted to find out the area of this circle or in other words you say that you want to find out the value of pi. And this is the same thing because this is equal to 1 unit and this is equal to 2 units.

So then that the area of this is equal to PI r square where r equal to 1 so it becomes equal to PI. So the area of the circle is equal to PI. So if you can calculate by using the MMC technique and we have thrown points at random. Let me show it so these are points at random thrown which are inside the circle which have fallen inside the circle. So these all these points that satisfy x I square plus y I square is less than equal to 1 okay.

So these are the points and then also consider this blue point say for example which have not been into the circle. So these points are not important points for us because these points do not contribute to the either the computation of Pi or you know the area of the circle. So for us the important ones or the important prize trials rather the important trials are the red ones. And so if there is a way for us to assign appropriate weight to some of the points that is the red points here and actually assign very little importance to the blue points.

Then we would do this thing much easier or they will be much more accurate if we know how to exclude these blue points from the red points, the red points being the important points, all right. So this is the main idea of this important sampling. So the methods that we have introduced so far, they generate arbitrary points from a distribution and these arbitrary points while we are saying because they are from their random points from a given distribution to approximate the integral.

So we have approximated the integrals by simply writing down that this is 1 over n sum over this f of x I and I equal to 1 to n n is the number of sampling points F x I is a value of the function at all those sampling points you sum all of them up divided by the number of sampling points this is equal to the integral. And basically what you have to do is that you have to integrate f of x. dx from some you know lower limit to some upper limit.

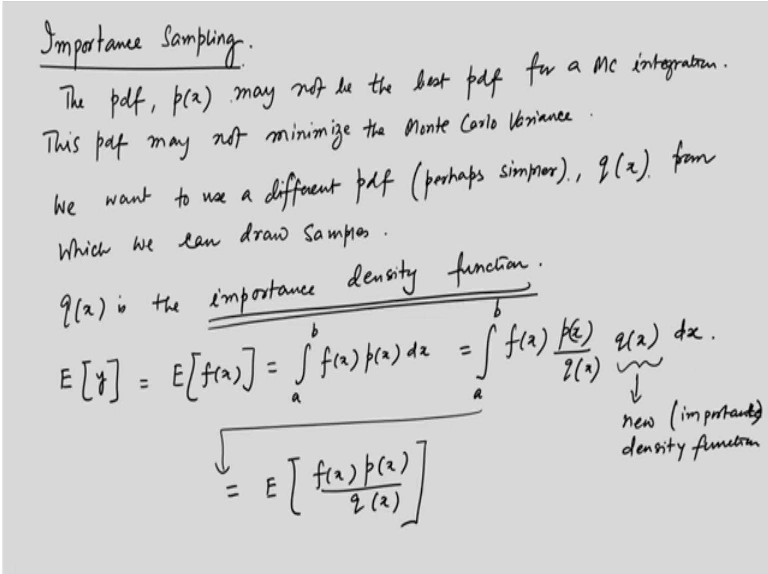
So this is what we have done. But so these are arbitrary points from a distribution to approximate the integrals okay and in many of the cases okay the points these points they correspond to the ones that are at the value of the function under consideration where the values are close to zero okay. So we are really talking of the blue points which do not contribute to the computation of Pi or the area of the circle.

So if the value of the function at those random points many of them would be there, if we are choosing at random without really looking at the nature of the function. Then, we will be wasting our computational resources and our time okay. So if the value of the functions are close to zero then those things will not be should not be considered in the sampling of the points in order to do the integral so because they contribute very little to the approximation to this approximation.

So in many cases the integral comes with a given density which is called as the probability density function or the pdf, okay. However, there will be cases where another distribution okay you carefully choose it that another distribution gives a better fit to the integral which means gives you a more accurate estimate of the integral. And we have to look for these another distribution and this another distribution is what is done in the important sampling.

So, this another distribution is more important is the important you know distribution which provides a better estimate of the integral. Let us see how to go about which will give an example which will make things clearer.

(Refer Slide Time: 19:53)



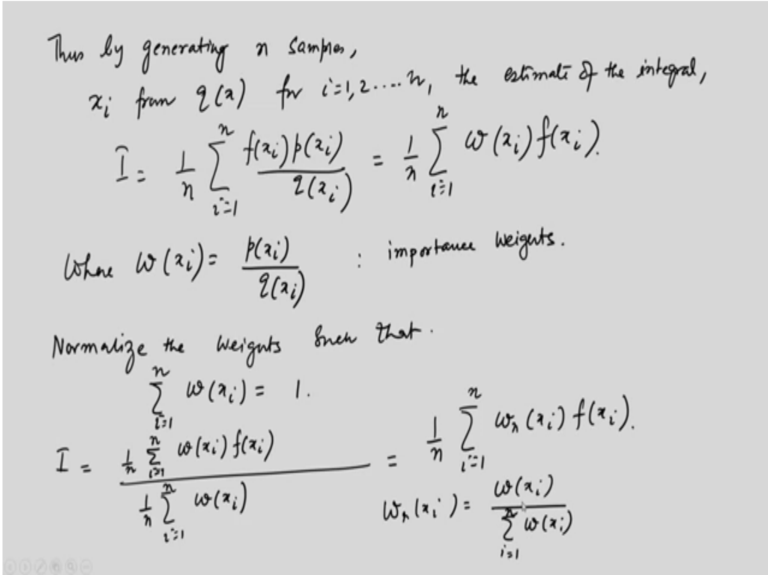
So what we have said is that the pdf, the probability density function p of x may not be the best pdf for a MC integration. So what it really means what is meant by a best is that so this this pdf may not minimize the Monte Carlo variance okay. And in which case we want to use a different pdf and perhaps simpler olit q x from which we can draw samples okay. So this q x is called as the importance density or importance density function.

So this is an important thing and how do we find q x is a matter of you know choice and one has to really understand the behaviour or the properties of the function that we want to integrate. We will see that so we can write E of y or the expected value of the function which is E as f x. So that is we know that between two limits lower and upper limits a and b is written as f x p x d x. So p x is our, the probability density function.

So this is equal to we simply rearrange things here so we write it as a p x and q x and q x and d x okay. So if you see this if you now call this as your new and in bracket which is we believe that this is the important or density function or importance as we said earlier density function then this is nothing but so this is equal to the expected value of f x p x / q x and so on okay and which is weighted by this q x.

So we have simply introduce this q x distribution and these it becomes the expected value of that of this quantity which is f x p x by q x which is weighted by the q x.

(Refer Slide Time: 24:02)

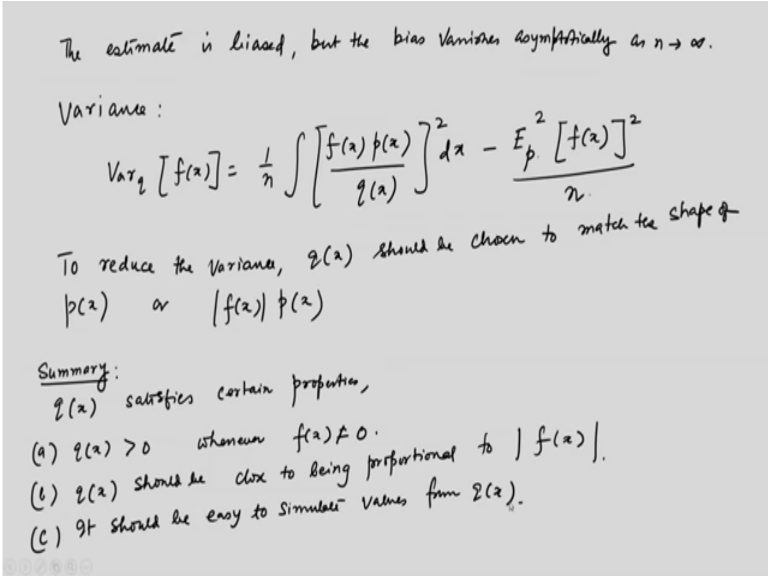


So thus by generating n samples which are from this xi from from qx for i is equal to 1 to n. And the estimate of the integral becomes i equal to 1 over n. sum over i equal to 1 to n f of x i P of x i divided by q of x i and this is called as or written as i equal to 1 to n. And we write W which is called as the importance weights it is written with a small w and f of x i okay and where w of well you let make sure that this is written as so this w of x I that is a smaller w so this is p x i divided by q x i okay.

This usually it is written like this and so this is the weight or which are called as the, these are the importance weights okay. So this is a ratio of the old probability distribution which is uniform or which is you know taken from another distribution. And q x is a new distribution so it is the ratio of that. And so we have to normalize the weights such that some over i equal to 1 to n. w x i should become equal to 1 and in which case we can write down this integral as simply 1 over n sum over i equal to 1 to n. I hope this is visible it is i equal to 1 to n. Let me then rewrite it more carefully so I equal to 1 to n and you have a w x i f x i.

So now instead of weighting it by the p which is a probability density function earlier one, we now weight it by the importance weights. And there is this of course has to be normalized I equal to 1 to n and w x I and so on. And this is equal to 1 over n i equal to 1 to n and w n x i f x i and so on ok where your w n x i equal to w x i divided by the normalized wxi i from 1 to n. Now this is the main issue that is the main concept that is involved here is that we have instead of this p x we have w x and w x is derived from this p x with another distribution q x which seemed to fit this better. So let us see a little more on the variance and so on.

(Refer Slide Time: 28:31)

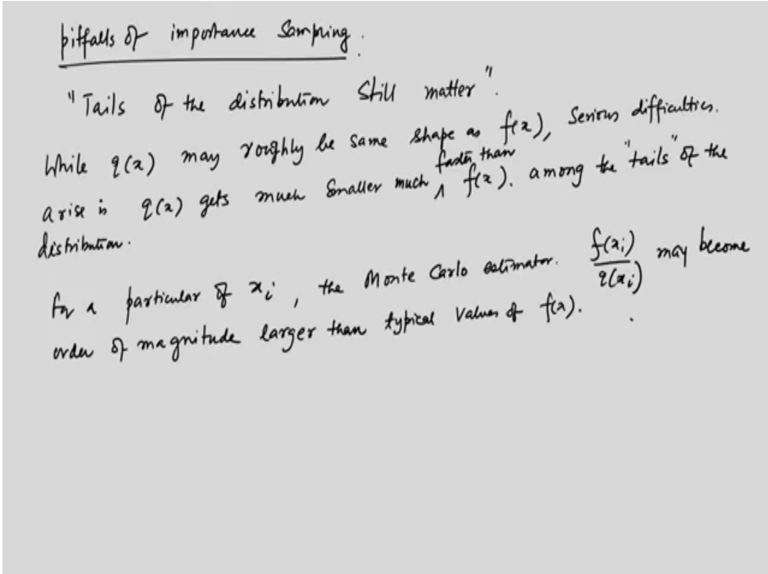


Of course the estimate is biased because instead of a uniform one you have taken a biased distribution which fits your problem more closely but this bias actually vanishes in the asymptotic limit. So it asymptotically as n tends to infinity okay. So the variance of this thing is given by variance of this q distribution is given by it is equal to 1 by n f of x p x square sorry this is divided by q x and then one has to square this minus d x.

This is square d x minus E for the p 1 this is for the p, pdf the original pdf square. So to say f of x this is that original thing that we have discussed. So this variance is for the new one minus the old one and to reduce the variance q x should be chosen to match the shape of p x or f x p x. So what I am trying to say is that q x is often is proportional to p x. So in summary a good important sampling q x satisfy certain properties.

q x is greater than 0 whenever f of x is not equal to 0 okay. q x should be close to being proportional to which is what I just said to f x and then of course what is important is that it should be easy to simulate values from q x okay it should be as I said earlier that it should perhaps be easy to simulate values so that we get a better estimate and things like that for the for the integral that we are trying to compute.

(Refer Slide Time: 32:31)



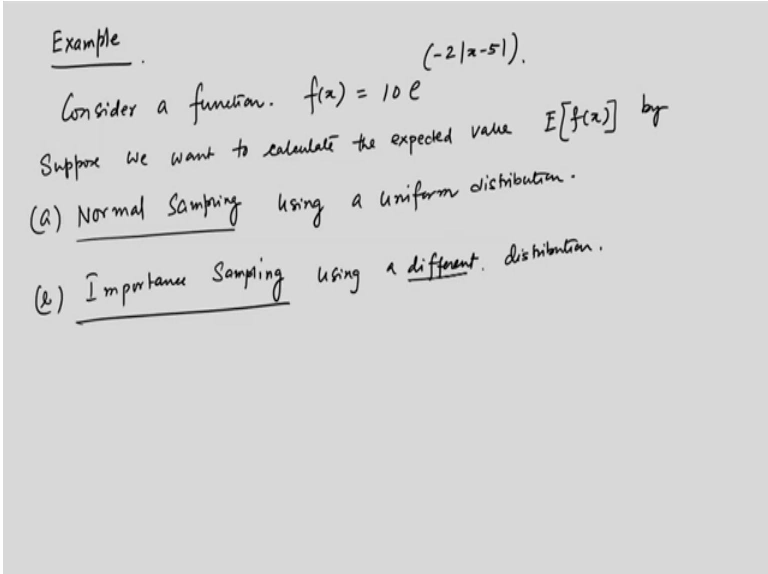
So this is just one thing that one wants to take care of so it says the let us say the final comment or rather this pitfalls of importance sampling because we are taking a ratio you need to be careful about the tails of the distribution of this distribution of q suppose q has a distribution which is you know having a normal or Gaussian distribution one needs to worry about the tails. Even though it is a sort of a slightly improbable situation that it would happen but the tails of the distribution, the new distribution that is the importance distribution q still matter okay.

So how does it matter? so while you know q x may roughly be same shape as f of x serious difficulties arise if q x gets much smaller than f x and why is it a difficulty because this importance weights are proportional to the ratio of these. I mean get smaller than f of x is not the right this thing gets smaller much faster than, much faster than f of x and around or among the tails of the distribution.

Because in this case what happens is, that for a particular value of x I, the Monte Carlo estimator which is f x i by q x i so maybe actually maybe come order of magnitude larger, than typical values of this okay. Rather typical values of q of f x not q f of x okay. So these are some of the things. So basically what I mean to say is that it is very highly improbable that you know in the tails of the distribution things, can these ratio can become much larger than the typical values.

Then they would start contributing dominantly which is what is not you know wanted. So but these are very improbable situations.

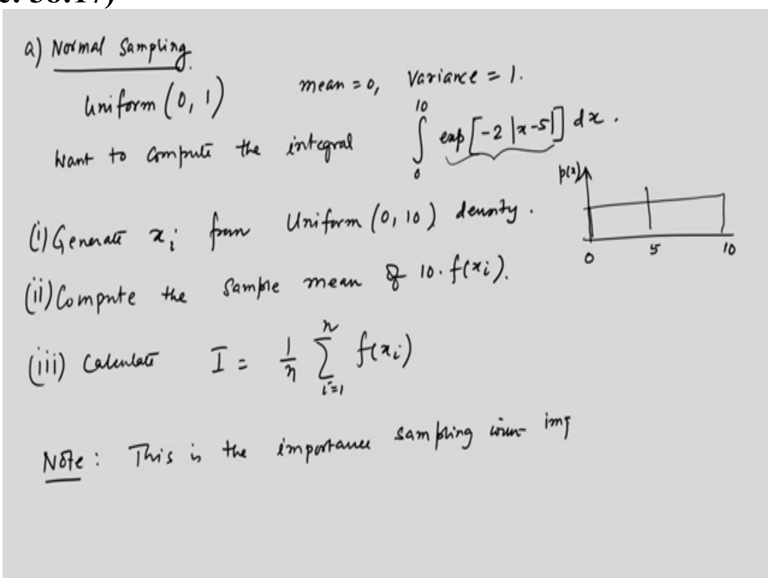
(Refer Slide Time: 36:28)



Let us see an example of how to use this important sampling and how important sampling can improve the approximation for a given integral. So consider a function f of x equal to 10 exponential minus 2x-5 okay. And suppose we want to calculate the expected value e f x by two methods. One is let us call it as a normal sampling which is what we have been exposed to from the beginning that is using a uniform distribution.

Or B the important sampling using a different distribution and how do we choose the different distribution, we will just come to that okay. So let us look at this normal sampling a all right .

(Refer Slide Time: 38:17)

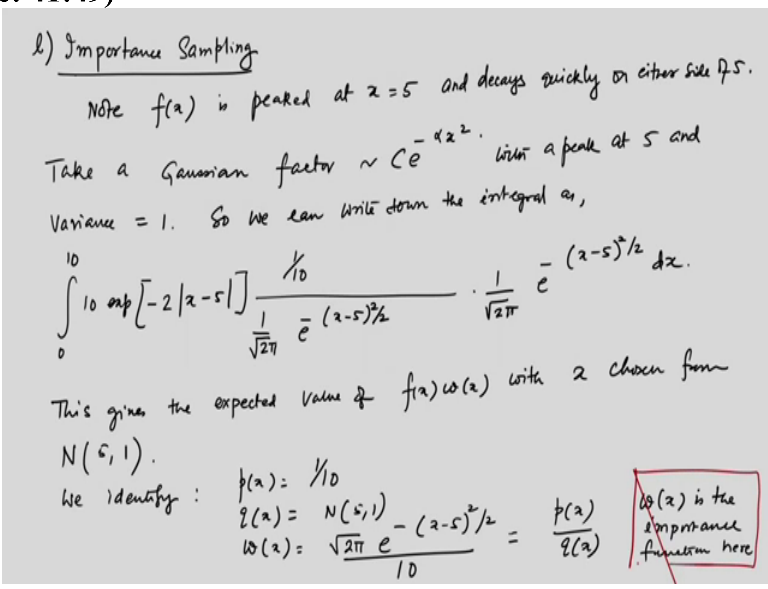


So we will use a uniform distribution let us write that uniform distribution as 0 & 1 which means the mean at 0 and the variance is 1. So let us write that mean equal to 0 and variance equal to 1 so we want to compute the integral 0 to ten exponential minus 2x mod x - 5 d x okay. So a regular approach would be to generate x i from uniform (0,1) distribution here because we have taken it to (0, 10) because this integral has an approximate value of 1, we could do that okay.

So this intake this integral has a value which is close to 1. S so this is this uniform density or a uniform density function. Compute so this is 1 and this is compute the sample mean, mean of 10 times f of x i okay. So basically you pick random numbers from this distribution uniform distribution which is like this ok. So this is from 0 to 10 and this is 5 say just to mark this. So this is your p of x. And so one chooses value values at random from this and then calculate i equal to 1 over n i equal to 1 to n f of x i to calculate the integral. So this will give the value of the integral.

So note one thing which will become more clear that the this is the important sampling with importance function so, w x is equal to nothing but your p x okay. So, now let us look at the more important thing which is called as a important sampling all right.

(Refer Slide Time: 41:49)



So for doing that take a note that f x is peaked at x equal to 5 and decays quickly on either side of 5 alright. See if that is the case doing or taking a uniform distribution is probably not such a good idea. So what we can do is that one can take a Gaussian factor which is like exponential minus you know some alpha x square etcetera with a peak at 5 at 5 and minimal variance and the variance can be equal to say 1.

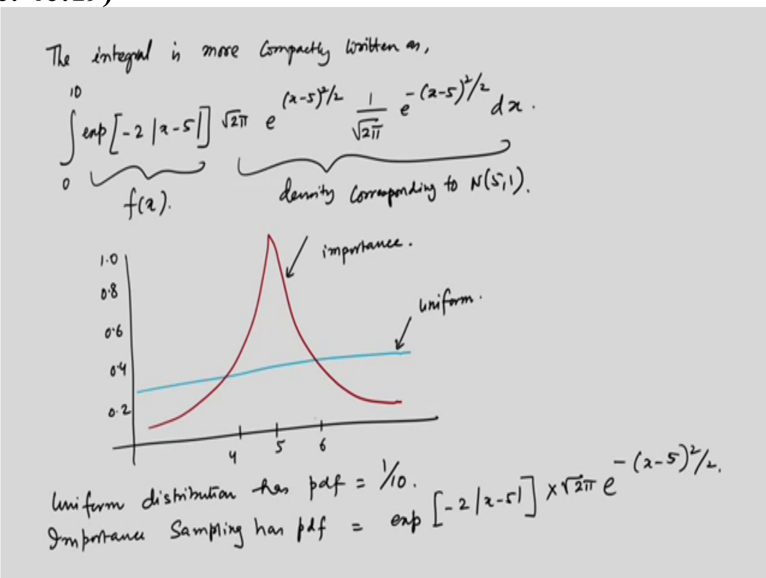
So we can write down the integral as so it is from 0 to 10 and we have this function which is minus five and then there is a 1 over ten and now there is a 1 over 2pi exponential minus x - 5 whole square by 2 and in 1 over 2pi exponential minus x - 5 whole square over 2 and dx. See exactly we have introduced a queue below and a queue above, so this is the importance weight.

So, this gives the expected value of f of x w of x with x chosen from not from a uniform distribution but from a normal distribution with mean at five and variance equal to 1.

Remember that we have used this language or terminology for denoting a normal distribution with the first of the argument that represents the mean and the second 1 represents the variance. So here we identify p of x equal to 1 over 10. q of x is equal to the normal distribution which is (5,1) and w of x which is equal to a root over 2 pi exponential minus x - 5 whole square by 2/10 and this is nothing but equal to p x over 1 x okay.

So this, please go back and see the definition this is exactly what we have done. So instead of a constant distribution which is like 1 over 10 and we have now taken a distribution which is normal with variance and a mean that is given here okay. So w x is the importance function here, let me write that here itself or you can call it the importance weight and so on. So let me box this.

(Refer Slide Time: 46:19)



So if that is the case then the integral is more compactly written as 0 to 10 exponential of minus 2mod x - 5 root over 2 pi exponential x - 5 whole square by 2 and 1 by root over 2 pi exponential minus x - 5 whole square by 2 d x okay. So this is the original f x and this is the density being integrated or so. This density corresponding to the and N(5,1) the normal distribution.

So let us plot this thing to have a more intuitive understanding. So this plot of the integrant from a and b. So let us take so let us take a different color for so. This is the uniform distribution all throughout which is same so this and this is equal to the distribution okay. And so we are showing it I mean it is slightly the it should have been slightly more rounded. But what is important here is the following.

So you have so this is like 2 and a 4 and maybe yeah well let me just do it. So this is 5 okay, so this is 6 and this is a 4 and so on. And this is of course this is like a point to 0.4 0.6 these are normalized 0.8 and 1. And so this is corresponding to the uniform the normal sampling and this corresponds to the importance sampling. You can see that this important sampling is likely to give better results because this is peaked at 5 where the function peaks it.

So you have another distribution from which you want to take and this distribution is a better than the uniform distribution. So you see that the value of the function falls off quickly on either side of 5 the uniform distribution does not know anything about it whereas the importance distribution that knows that this is you know it falls off quickly and so on. So the uniform distribution has pdf equal to 1 over 10.

And the important sampling equal to this exponential minus 2x - 5 into root over 2 pi exponential x - 5 whole square over 2 okay. So, this is it gives you an intuitive way to understand why the importance sampling gives much more accurate results for this. It is a very important thing in the study of Monte Carlo where we should always weight it by an importance function rather than by doing a normal sampling especially for complicated integrals. This makes much more you know or rather provides much better estimate of the integral.