Metropolis-Hastings

Now, many times, even sampling from the conditional distributions is not possible. And in these circumstances, we need to resort to the second main algorithm using, used in Bayesian computations, which is the Metropolis Hastings algorithm. Essentially, the Metropolis Hastings algorithm is used when we cannot even sample draws from this conditional posterior densities. But we can, however, sample from another distribution, say g. Right. In in those cases, we are going to resort to this second main algorithm. Specifically, in a nutshell, what we do in the Metropolis Hastings algorithm is a protocol, essentially, we follow a protocol with the following steps

We start again, with initial value of the parameters, and then we sample a candidate value of the parameter from a density g from which it's easy to sample. We then evaluate as in seen in equation 32, we compute the ratio, which is essentially a ratio of the posterior distribution of the parameter evaluated at a candidate value of the parameter and then the previous value of the parameter multiplied by again a ratio of the function that can be proposed what is called the proposal density g evaluated again, in this case, the denominator at the previous value as a parameter and in the denominator carried value within sample another value u from a uniform distribution, and ultimately compare that value of u against the ratio, and we accept that can be it or not, depending on whether the ratio is larger than the value of u or lower than the value of u. And we do this again many many times. Right.

So the idea, the key component here is this ratio, which is element in equation 32, which is the third step of the Metropolis Hastings algorithm. And this ratio has, as I said, two components one is the ratio of the posterior distribution of the parameter from which we cannot sample. But however, we can still evaluate it both at the candidate value of the value of data sample, sampled from the proposal density, divided by the value of this posterior distribution evaluated at the previous value of the parameter. So the parameter we sampled at the previous iteration of the algorithm. So essentially, the larger this ratio, the to posterior density or procedure distributions, the higher the likelihood that the candidate is actually a good candidate, and we correct this ratio, or we multiply this ratio by the second component, which is the ratio of the proposal density evaluated at the previous value of the parameter, the candidate value, this is sort of an adjustment to account for the fact that some proposal densities might lead might disproportionally to some values of the parameter rather than others.

Now, the question is, well, what what is a good g?, what is a good proposal density? What is a good density from which we can sample candidates? And here the Bayesian literature is very extensive. However, something or a couple of proposal densities that have worked very well for me in the applied work are random walk normal and random walk uniform proposals. Essentially, we're working with a normal random work proposal, what we do is we draw a value, a candidate value for theta from a normal distribution with mean the previous value of the parameter and with a certain bias, or we can draw the proposal density from a uniform distribution, it's centered around the previous value of the parameter.

And I go into these in much more detail in the slides but the point is that by using similar symmetric, symmetric proposals, such as a normal uniform, the computation of the ratio becomes simpler and essentially just reduces to a ratio of the posterior densities evaluated at the previous value of the of the back of the parameter and at the candidate value. The overall point is using symmetric proposals such as a normal was a uniform helps simplify this process.

Let me give you an example of when one would use Metropolis Hastings algorithm, rather than a Gibbs sampling, I suppose. In this exercise, we're trying to draw inferences about the correlation coefficient followed by a variate normal distribution, I have a standard, bivariate normal, and we want to say something to draw inferences about the correlation coefficient. And, as usual, posterior is proportional to the likelihood times the prior and assume that we adopt uniform prior for rho a uniform between minus one and one which is the balance of correlation coefficients.

Now, the posterior distribution of rho given by this equation here, which as you can see, very easily we verify doesn't have a closed form, we cannot really sample from this distribution, we cannot also not use the Gibbs sampler. Now, only one parameter, how can we how can we condition these into a conditional distribution? So essentially, what we're going to do is we're going to use a Metropolis Hastings.

So again, what does it mean using a Metropolis Hastings? we need to come up with proposal density, we need to generate candidate values from that proposal density. And then compute the ratio and compare it to the value of a unit of variable from a uniform distribution. For instance, given that rho the correlation coefficient is bounded between minus one and one, simple proposal will be a uniform proposal density with very small increments. So, each iteration of the algorithm we start from the previous value of rho and sample a value from a uniform between minus point one and point one for instance. So this was to allow the sampler to explore the whole possible space of the parameter rho which is minus one.

So again, in our problem, the idea is we start with a certain value of rho for instance, zero. And we start at each iteration of the algorithm, we draw a candidate value for using a uniform proposal, we evaluate the posterior distribution of rho which we know just that we cannot generate values from these, but we know that once we have a value of a rho, we can collect the value of rho here and we can compute this value we evaluate this posterior distribution and the candidate value for rho and the previous value rho compared with a u drawn from a uniform distribution and then accept this candidate value as the new value of rho or not depending on whether the ratio is larger or smaller than u.

We have, you have an accompanying script called Metropolis Hastings algorithm for rho in r, which you can use to see how this actually works in practice, how these five steps actually work in practice. And once you do that, you obtain something like this we start drawing what is of rho, starting from zero, and very quickly, the values of rho stabalize around minus point six which is the value the true value of rho all in the exercise.

Now, in most cases, so this Gibbs sampling, and Metropolis Hastings algorithm are, as I mentioned, the two main algorithms used in modern Bayesian computation. For most or for many practical problems we're going to use both in estimation in estimation of any more. For instance, you have an example here in the slides, and also in the scripts of theoretical logit model in which you have N individual observations nested within, within J countries. So, equation 37 gives you the equation for a hierarchical logit model with country specific random effects drawn from a normal distribution with mean zero and some variance sigma square and doing computations, you can see that the regression coefficients are the betas have no close from the random effects, also, the posterior distribution had no close form. However, sigma squared, the variance of the random of his country and the effects they do have close form have an inverse gamma posterior distribution.

So here, when estimating this hierarchical logit model, one would combine two steps one would use, essentially. And here are the posterior distributions for you to take a look when working through this exercise, but the point is that our Bayesian estimation approach in this setting would combine Metropolis Hastings steps to 10 values for better and for the country specific random effect, and Gibbs sampler to draw a sample for sigma squared. And again, in many cases are many interesting problems, you will have to combine both Metropolis deep sampling and Metropolis Hastings steps, you also have another exercise to estimate the hierarchical probit rather than logit. And you will see that in the case of the hero go pro it everything can be done using Gibbs sampling, because in this case, you in the probit model we can break the problem of the problem of deriving joint posterior distributions into a series of conditional posterior distributions.