Gibbs Sampling

In the previous example, everyone, everything was very smooth in the sense that we could easily derive the posterior distribution of the parameter p, which turned out to be beta. And we could then apply very nice formulas to compute the posterior mean of this, this beta distribution, or to compute the probability that ultimately, these formulas come from solving certain integrals. So for instance, we could compute the posterior mean, because we arrived at the closed form. Equation nine, which results from solving this integral will also compute the probability that candidate A wins also coming from this integral, this is very nice.

Now, in many, I will tell you most interesting social science programmes we are A, you'll not be able to solve these integrals or not be able to summarise the posterior distribution using standard closed form known formulas. And here's where more than Bayesian computation methods come handy. And the idea is, well, when we are not able to solve these endeavours, what a Bayesian does is to resort to simulations, right.

And in particular, we resort to what it's known in more than basic computation of Markov Chain Monte Carlo simulation. In the slides, I give a little bit of detail about how, why this Markov chain Monte Carlo simulations are called the way they are. But essentially, the idea is that when I cannot come up with closed form solutions for the parameters of quantities, interests of the parameters, I can generate simulation.

And the two key and, they're kind of very different, that many different potential ways to simulate values of the parameters that are amenable to them, maybe some summarise or draw inferences on these parameters. In modern Bayesian computation, there are two key algorithms that we use. One is the Gibbs sampler, and the second one is Metropolis-Hastings . And the idea is we resort to this simulation techniques, and in a nutshell because we have models for which it's difficult or impossible to come up with close form solutions for the parameters when there are no known formulas we can use to obtain their posterior mean or median or variance.

However, you can use Monte Carlo simulation techniques, or code chain Monte Carlo simulations to that will enable us to draw inference about these parameters even when we don't have a closed form solutions. So let me tell you in a nutshell I what these two key approaches to Bayesian simulation are, these two main algorithms, the Gibbs sampler and Metropolis Hastings. What they do and essentially, the idea is the Gibbs sampler is the most basic algorithm used in modern Bayesian computation. And it is applied when I have a model with several parameters, think of any interesting model probit or logit. And experiment have a joint posterior distribution of all the parameters that's very difficult to draw samples from. But we can split this joint posterior distribution in a series of conditional posterior distributions. We can sample from this conditional posterior. This is kind of what the Gibbs sampler does, break down the problem of generating samples of some of the parameters from the simulation because they cannot simulate from the joint of series distribution. But if I break down the problem into a series of conditional distributions, I may then be able to simulate.The metropolis Hastings algorithm is a generalisation of the game sample that we are going to apply when even these approaches of breaking down the joint posterior distribution of the parameters into a series of simple conditional distribution doesn't.

So, let me illustrate each of these two key or main bayesian algorhithms during the basics of modern Bayesian computing, as I said to both to have a vector of parameters big theta and with the theta they have k parameters. So, suppose that the joint posterior distribution of the parameters has a very complicated form then there's no closed form solution for these parameters and I cannot even simulate from this through distribution however, I can break down the problem or split the joint posterior distribution into a series of conditionals. So, I can suppose, the conditional distribution of theta one given the rest of the parameters is something I can recognise. The same for a conditional distribution of the second parameter given all the other parameters under the data of course, and the same for the conditional posterior distribution of theta k given theta one theta two, if this each of these conditions are recognisable, or in other ways I can generate samples from these parameters from the from these distributions, there is a version of the law of large numbers, which I go into a little bit more detail in the slides, that tells me well, if this works, if I can break down this problem into a series of conditional distributions, I can sample from each of them and then I can approximate the joint distribution of all the parameters.

So, essentially, what I what the Gibbs sampler does is starting from some initial value of the parameters. If it had a problem with many parameters, I break down this joint posterior distribution from all the parameters into a series of conditional distributions, each of which is recognisable meaning, I can sample generally draws from each of this conditional and if I repeatedly sample from each of these conditional distributions, I will approximate the stationary distribution of the parameters, the equilibrium distribution of the parameters.

Let me give an example in practice. So, suppose to have simple random sample from a normal distribution, where I do not know the mean, and the variance sigma square of this distribution. And the goal of this exercise is to estimate mu and sigma squared, or to be Bayesian to derive the posterior theory, distribution of mu and sigma. And again, in this exercise I'm going to show you today, they can matter, mantra, the posterior is proportional to the likelihood times the prior.

Right. So, this is what equation 18 the posterior distribution of mu and sigma squared, even the data is proportional to the posterior distribution of the data times the prior distribution of the parameters. Okay, so, again, we're going to go through the four steps of Bayesian inference, but applied to these exercises.

Since we said that, we have a sample taken from an old distribution, the sampling distribution of the data is given by equation 19. So, the posterior distribution of the parameters is going to be like says, like, you can see in equation 20 a normal likelihood times the prior distribution of the parameters.

The second step, any bayesian analysis is coming up with the priors for the parameters, mu and sigma. I could generate a joint prior. So essentially generating prior distribution it's together for mu and sigma square. It is here typically, though, to use different priors for different sets of parameters in this case, to specify a prior for mu and a different prior for sigma square. So, we can choose what we mentioned when we called conjugate priors, priors such that the posteriors are going to be of the same family as in distribution many proiors. So, turns out that for the mean of a normal distribution, the conjugate priors are normal, for the variance of a normal distribution, the conjugate prior is the inverse gamma distribution, okay. So, point is, once I have that, once I choose a normal prior from you and an inverse gamma, like the one we see in equation 23 for sigma square, I can then move on to Step three, which is deriving the posterior distribution of the parameters.

Okay, so, let's say that I have the data, likelihood, I have the prior for mu have the prior for sigma square and applying the Bayesian mantra posterior proportional to data times priors, I arrived at equation 24. The problem there is that equation 24 has, is not something I can recognise . It's an equation that doesn't have a very clear cut or close form solution. So in other ways, I cannot analytically solve this equation, it's also impossible to simulate from this equation. And this is however, when the Gibbs sampler comes in handy, because what does the Gibbs sampler do?

Well, we said, break down this problem of deriving the joint posterior distribution of the parameters into a series of conditional distribution. And our way of saying it in this example is, when I want to derive the conditional posterior distribution of mu, let's take sigma squared the other parameter as a constant, and when I want to derive the posterior distribution of sigma square, I will take mu as constant and then I just repeatedly sample from these conditional distributions. By doing so, I approximate their joint posterior distribution. So, to derive their conditional posterior distribution of mu what I do is take sigma square, I have two parameters in this exercise. So, this is a special example of the equation. The more general formulation of the Gibbs sampler has k parameters, in this particular case I have two parameters mu and sigma squared, and what I do is either add the conditional distribution of mu taking sigma square scaling. So, in other words, take any term in this joint posterior distribution, that doesn't depend on mu as a constant.

Essentially, I can forget for a moment about all the terms that do not depend on mu and with a little bit of algebra, I arrived at the fact that again, this is the slides go into much more detail about this, but the point is, if I take say sigma square constant, when I derived the posterior distribution of mu, I know a little bit of algebra, when I arrived at his equation like 26, where the conditional posterior distribution mu resembles a normal distribution. Equation 26 looks like the kernel of a normal distribution with certain mean certain words right.

Is that normal will mean given any in equation 28 and variables given by equation 29 and this, I know how to generate values from a normal, I have, I know how to compute means banishment normally I can simulate very easily in R using the command rnorm to generate values from a normal distribution.

Similarly, when either the conditional posterior distribution of sigma square, I treat all the terms that do not depend on sigma square as constant, and, again doing a little bit of algebra, but i arrived is at question 31 one which is the kernel of an inverse gamma distribution. Again I do know how to generate samples from an inverse gamma using a command such as rgamma in r. The point is that starting from a joint posterior distribution that was very complicated, let me go back and show you this was the joint posterior distribution. I had no idea what this is, is this a normal? is this is norm?, is this is an inverse gamma? no this is a complicated equation. I have no idea how to compute posterior means. assume variances, how to summarise the parameters right. So, what I did is I broke down this problem of generating of computing, deriving the joint posterior distribution of the parameters into a series of conditional distributions. And once I did that, I arrived at the problem in which I end up with a normal distribution and an inverse gamma distribution and those I can think, you know, how to work with right.

So, essentially, the way in which the Gibbs sampler will work in my problem would be start with initial values for sigma square mu, choose norm a normal prior for sigma, for mu sorry, and inverse gamma prior for sigma square and then repeatedly draw samples for mu either one normal conditional distribution, for sigma square an inverse gamma distribution and that will approximate the joint posterior distribution of two parameters.

You have a script, comes together with this slide called Gibbs sampling normal distribution in which essentially I show you how to, starting from initial values, the parameters sigma square and mu, I start drawing sample and as you can see, I start drawing for instance in this figure, what I see is that I can start generating samples for mu out of a normal distribution Originally, the values of the normal kind of all over the place, but pretty quickly they stabilise around a value of two. Wn reach their equilibrium. So, the posterior mean the value mu is going to be centre around two. Similarly, for sigma square I started generating samples for an inverse gamma at the beginning that samples during the [inaudible] but eventually they will stabilise around the value close to four.

And this is kind of in a nutshell what the Gibbs sampling does. Break down a complicated problem into a series of issue conditional problems. And, and they're very generally conditions that hold under most, for most social science problem, doing this trick of simulating one, the value of one parameter taking the value of the other parameter is given going back and doing this many times.

Ultimately, leads me to the stationary distribution, the posterior distribution of more experiments, essentially, I reached the approximate the joint conditional posterior all the parameters by doing this trick of breaking down the problem into a series of conditions.