Franzen (2008) describes the Gibbs sampler which is a particular MCMC algorithm working with conditional states. Each iteration of the Gibbs sampler cycles through the conditional distributions of all the parameters. In each iterative step, new parameters are generated and the conditional distributions are updated for the next iteration. It is suitable in situations where the joint distribution of the parameters of interest, is difficult to calculate, but the conditional distributions are possible to simulate from. This iterative procedure makes the process approach the equilibrium (Franzen 2008). The posteriors of the parameters in the mixture model of formula (1.5) are estimated with the Gibbs sampler algorithm. The posterior distributions for all parameters, generated from the prior and likelihood distributions, are expressed conditional on one or more of the other model parameters (Franzen 2008). Gibbs sampling algorithm has the following properties: it is a particular MCMC algorithm which samples from the conditional distribution while other parameters are fixed and updates a single parameter at a time. Chai *et. all* (2016) explains that Gibbs sampling starts with a random setting of hidden states and then updates each hidden state according to the probability distribution conditioned on all the other states and the fixed parameters. The procedure of Gibbs sampling can be described as follows (Chai *et. all* 2016).

* suppose that *}* are the unobserved random variables which states are hidden and *}* are the observed variables which states are fixed. Then the variable vector };
* start a random state for vector in which the state of is fixed and the state of is random. Then, we can make the data sampling for one of variable in , for example, , according to the probability distribution *P( | ,Y).* *In P( | ,Y), ∈ X \ { }.* In Gibbs sampling, *P(|,Y)* is also called as transition probability for ;
* next, depending on the result of sampling  *=* , we make the data sampling for another variable according to the transition probability for *P( | ,= ,Y)*. This step is iterated until the unobserved random variables in are all sampled. Then a piece of complete data sample for is obtained.

If *}* is the random variable vector for a BN, it should obey the conditional independence assumptions. For node , it is conditionally independent to any other nodes given the value of Markov blanket of . Then, we can simplify the calculation of transition probability, for example, *P( | , Y ): P( | , Y )= P( |*  ), where denotes the value of Markov blanket of (Chai *et. all* 2016). It is obvious that Gibbs sampling is very suitable for BNs. We can utilize Gibbs sampling to improve EM algorithm for BNs parameters estimation (Chai *et. all* 2016). If such sampling is performed, these important facts hold:

* the samples approximate the joint distribution of all variables,
* the marginal distribution of any subset of variables can be approximated by simply considering the samples for that subset of variables, ignoring the rest,
* the [expected value](https://en.wikipedia.org/wiki/Expected_value) of any variable can be approximated by averaging over all the samples.

In summary, Gibbs sampling is technique for generating random samples of multivariate data when you have limited information about the joint distribution. In Bayesian updating without a prior conjugate, the Gibbs update on a randomly chosen subset of the new full data set may be performed since previous data points are dependent on the new data. The size of the subset may be adjusted to achieve an appropriate trade-off between speed and accuracy. The data are generated by the following process: first, is sampled, and then the observables from a distribution which depends on are sampled, i.e., . Note that can take a variety of parametric forms. Thus, for classical Bayesian inference, assumption about a likelihood function for the data must be done, and then parameters may be estimated. However, here every sample from the posterior distribution is some setting of “guess parameters”. Posterior means can be computed to get a single “best” value of parameters (averaging the samples). The “best” in that case would be in the sense of minimizing the expected squared distance from the true parameters (r/learnmath 2012).

A collapsed Gibbs sampler integrates out ([marginalizes over](https://en.wikipedia.org/wiki/Marginal_distribution)) one or more variables when sampling for some other variable. For example, imagine that a model consists of three variables , , and . A simple Gibbs sampler would sample from , then , then . A collapsed Gibbs sampler might replace the sampling step for with a sample taken from the marginal distribution , with variable integrated out in this case. Alternatively, variable could be collapsed out entirely, alternately sampling from and and not sampling over at all. The distribution over a variable that arises when collapsing a parent variable is called a [compound distribution](https://en.wikipedia.org/wiki/Compound_distribution); sampling from this distribution is generally tractable when is the [conjugate prior](https://en.wikipedia.org/wiki/Conjugate_prior) for , particularly when and are members of the [exponential family](https://en.wikipedia.org/wiki/Exponential_family) (Liu 1994).