The generative model is trained with a large amount of data. After training, it is able to generate data similar to the initial set of data. As it has been mentioned in the sub-sections “*Mixture model*” and “*Baysesian mixture model (Bayesian estimation (inference) for(on) mixture models via posterior simulations*“ (section 1), there are some methods developed for pre-training for generative modelling: Maximum Likelihood Estimation, variance algorithms, Markov Monte Carlo method, and others (Lu *et. al.* 2019). That is, the generation can be done by sampling from prior and weighting by likelihood (it is the expectation maximization method), using some guide distribution (it is called importance sampling), using Markov Monte Carlo (iteratively sampling the unknown variables of the model from their conditional distributions), or variational methods. Markov chain sampling methods for Dirichlet process mixture models are thoroughly explained by Neal (2000). In the BCM generative story, clusters are generated first and each data point is repesented by distribution over clusters. Prototype is generated by sampling uniformly over all observations (Kim 2015), i.e., initially it is presumed that every cluster is equally probable. Each element of the feature indicator vector that indicates important features for the cluster is generated according to a Bernoulli distribution with hyperparameter (Kim 2015). The distribution of feature outcomes for cluster s is generated so that it mostly takes outcomes from the prototype for the important dimensions of the cluster. It is determined by vector g indexed by possible outcomes through:

|  |  |
| --- | --- |
| (1.16) |  |

Here, and are constant hyperparameters that indicate how much we will copy the prototype in order to generate the observations (Kim 2015).