

1. The first step is finding the probability of model, there are multiple distributions (the mixture distribution) in the mixture model. Then, we need to find prior of parameters which in each multiple distributions. Then finding the likelihood which is $P(X|\theta)$, the probability of observed data given by parameters. Then getting the Bayesian Inference for posterior distribution of parameters $P(\theta|X)$. we can use MCMC, HMC etc in this step. Then sampling the new data from the posterior distribution. In mixture model, we will sample a data point through one of the mixture distribution and the distribution from the mixture distribution. Then, we repeat sampling many times to get the new dataset.

2. first, we need to finding the probability of model including the observed data X and parameters θ . Then calculating the posterior distribution by observed data and prior of parameters. Then we get the new data points given X , predicting by distribution of them, we can get the $P(\bar{X}|\theta)$ which is the likelihood of new data given parameters. we can calculate posterior predictive distribution by $P(\bar{X}|X) = \int P(\bar{X}|\theta) P(\theta|X) d\theta$. Then we can sampling from the posterior predictive distribution, and predict by new data points.

3. There are three Assumption: MCAR, MAR, MNAR.

First, we need to get the probability of model which creating the regression model for $Y|X$, and then get the parameter of missing value. Then getting the

prior for parameters. Then we can find the joint distribution of regression and missing value (model). By MCMC, estimate the missing value with exact entries. and sampling from the joint distribution. we can check the parameter fits situation with regression models and X . we can separate to individual regression model. then we can impute the missing values by the posterior ~~to~~ predictive distribution. plotting the graph and the model we choose will be based on the assumption. Then using the values which impute to do the regression analysis.

4. For Final project, we finish choice of dataset and EPA part.