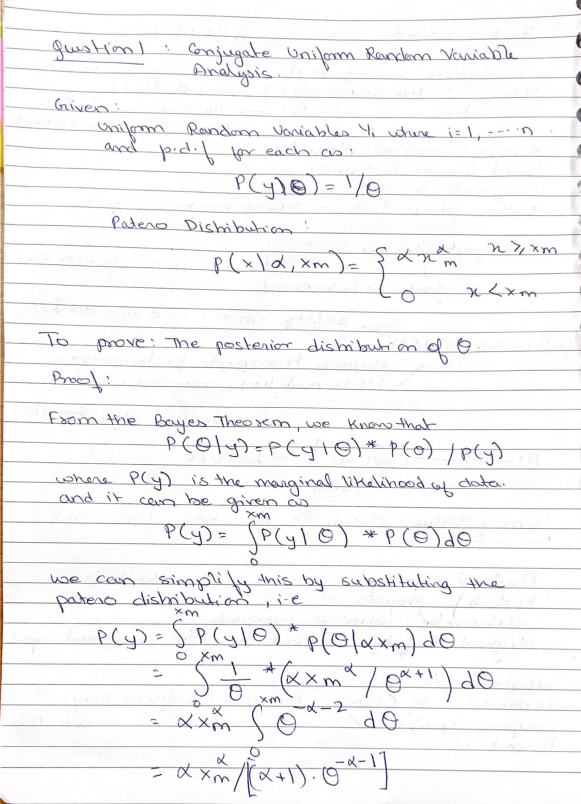
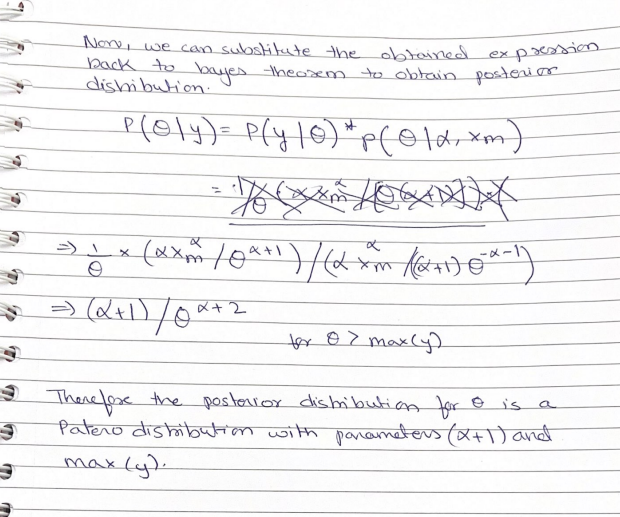
**Statistics Assignment 3**

Name: **Danita Anubhuti Prakash**

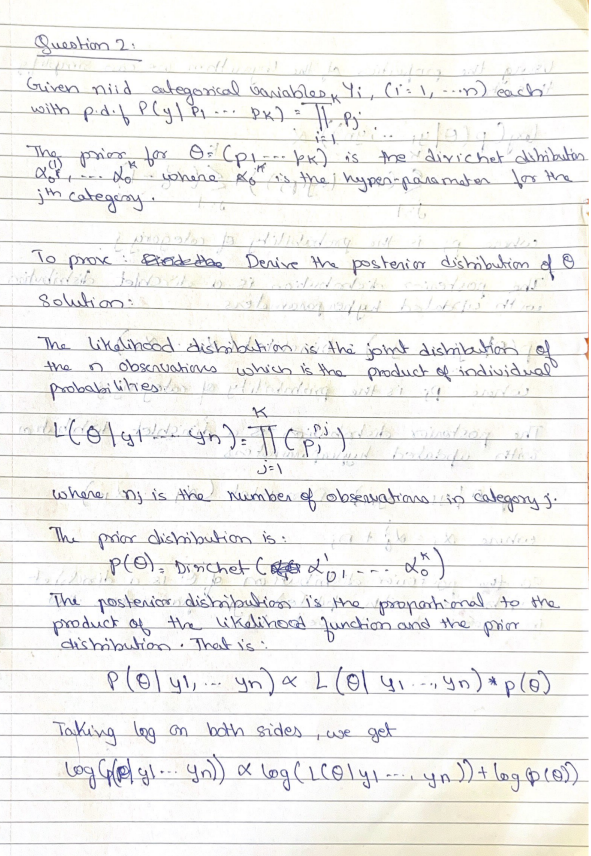
studentID**: s4745511**

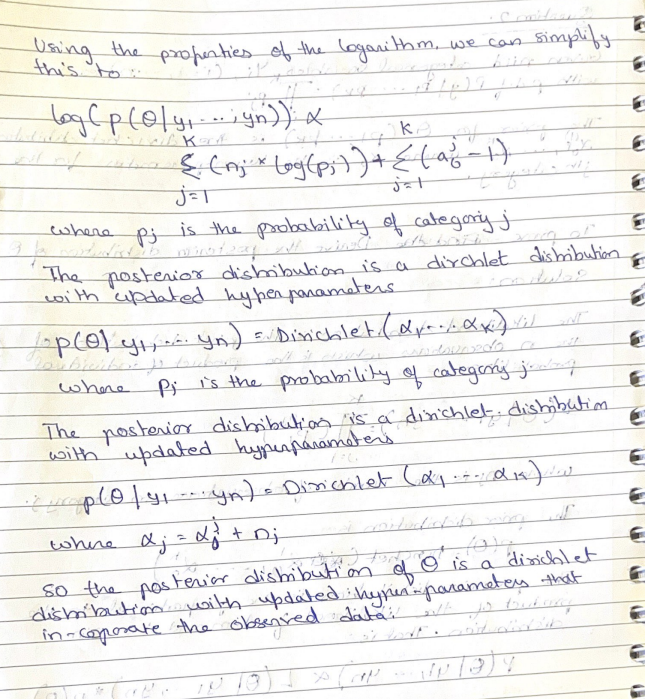
**Question 1:**

****

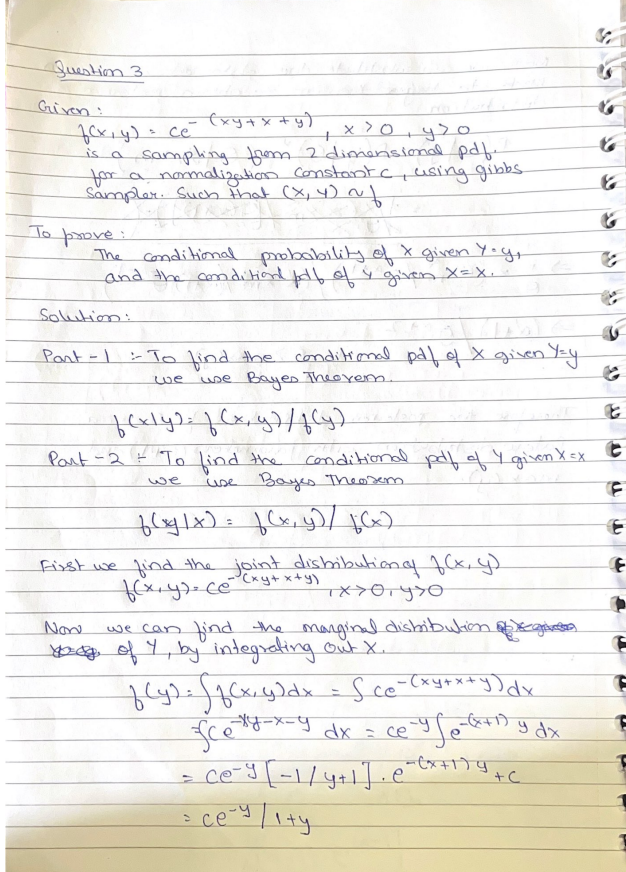
****

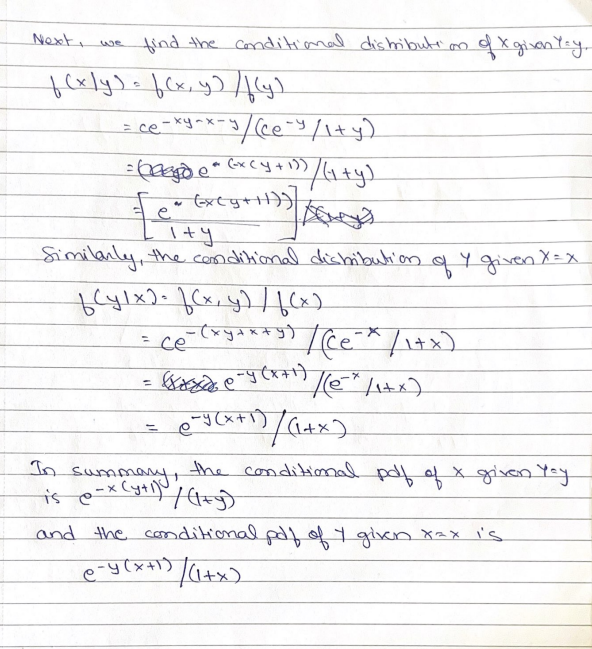
**Question 2:**

****

****

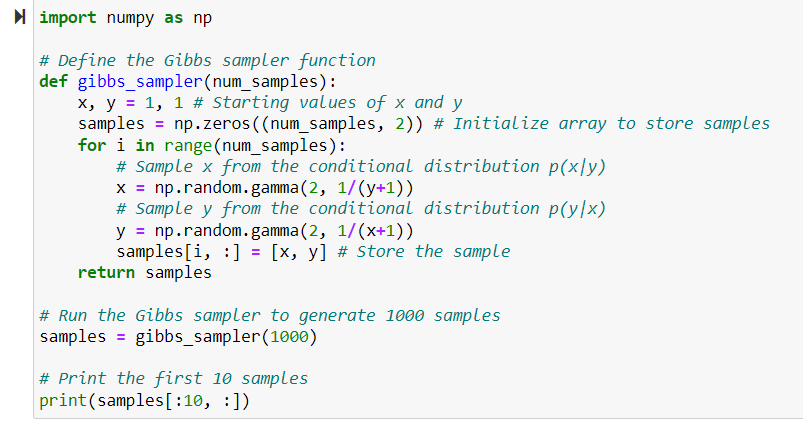
**Question 3 a:**

****

****

**Question 3 b:**

Consider a sampling from the 2-dimensional pdf f(x, y) = c e ^ (−(xy+x+y)), x >= 0, y >= 0, for some normalization constant c, using a Gibbs sampler. Let (X, Y ) ∼ f. The working code that implements the Gibbs sampler and outputs 1000 points that are approximately distributed according to f is as follows:

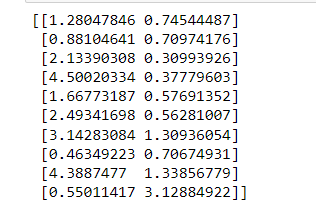
****

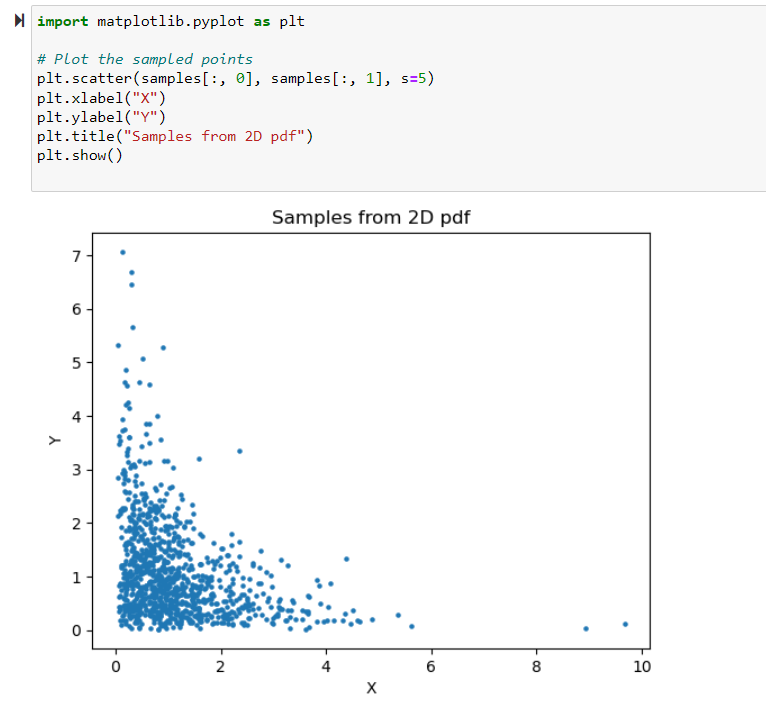
Explanation:

Here, f(x, y) is the joint pdf that we are trying to sample from. The other functions gibbs\_sampler(num\_samples) decide the conditional pdfs for x and y respectivly by taking the input of number of samples to be considered. After initializing the values with x and y as 1.0 iterations are run based on number of samples. Here Gibbs sampler for 1000 iterations are used. The following is the output of the some of samples.

In each iteraton, a new value of x and y is sampled from the function np.random.gamma(2, 1/(y+1)) and np.random.gamma(2, 1/(x+1)). This is then stored in the samples array and returned. Here we do not use the f(x, y) function, but it is implicitly used by the gibbs distibution.

OUTPUT:





GRAPH EXPLANATION:

It appears that the samples are concentrated in the lower left corner of the plot, with a central region of higher density and upper and lower regions of sparser density. This is in line with the target distribution's behaviour, having a larger probability density in the lower left corner and a decreasing probability density towards the upper and lower portions of the plot.

**Question 4 a:**

Model 1 assumes a binomial likelihood function and uses a sigmoid function to map the linear predictor to a probability. The model has two parameters alpha – the intercept and beta for the coefficient.

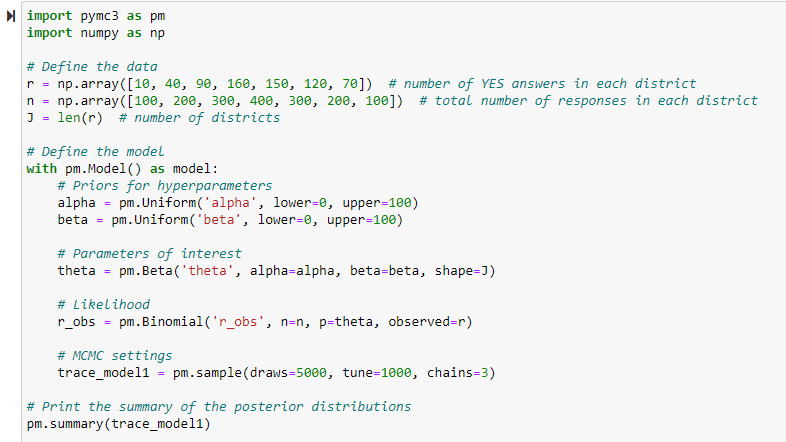
Prior: alpha ~ Uniform(0, 100), beta ~ Uniform(0, 100)

Likelihood: y ~ Binomial(n, sigmoid(alpha + beta\*x))

From the data given,

r = Number of yes answers in each district, n = Total number of respondents in each district, j= number of districts

Code:

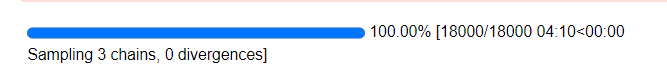


Explanation:

it uses uniform priors for the hyperparameters alpha and beta, and a beta distribution for the parameter theta\_j for each district. The likelihood is defined as a binomial distribution with the observed number of YES answers and the total number of responses in each district.

The pm.sample function runs the MCMC algorithm with 3 chains, each with 5000 draws and a burn-in period of 1000 iterations. The pm.summary function prints the summary statistics of the posterior distributions for each parameter.

From the code we try to sample 3 chains with zero divergance.



**Question 4 b:**

Based on the distance from the test centre and the student's level of preparation, this Bayesian linear regression model forecasts the number of right answers a student will provide on a test. The predictors are mapped to a mean response using a linear function under the assumption of a normal likelihood function. Beta0 (the intercept), beta1 (the coefficient for distance), and sigma (the standard deviation of the errors) are the three parameters that make up the model.

Prior:

beta0 ~ Uniform(-10, 10),

beta1 ~ Uniform(-10, 10),

sigma ~ Uniform(0, 100)

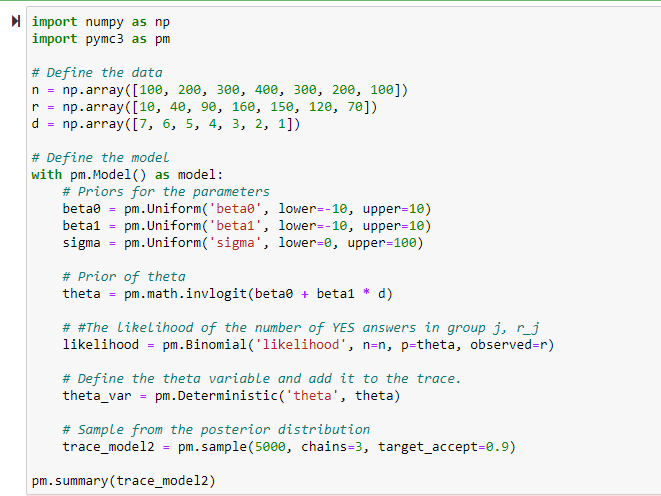
Likelihood:

num\_yes ~ Binomial(num\_answers, sigmoid(beta0 + beta1\*distance))

From the data given,

r = Number of yes answers in each district, n = Total number of respondents in each district, j= number of districts

Code:



Explanation:

The goal of the code is to estimate the probability of YES answers- theta as a function of the predictor variables. The prior distribution for theta is specified as the inverse logit transformation of beta0 + beta1\*d. The likelihood function is specified using the Binomial distribution with the observed data (r and n) and the probability parameter (theta). A deterministic variable is created for theta (theta\_var) using the pm.Deterministic functionmallows the user to calculate and store the value of a variable that is determined by other variables in the model. Finally, the model is fit using MCMC sampling algorithm, and the posterior distributions of the parameters are summarized using pm.summary. The target acceptance rate is set to 0.9 to ensure good mixing and convergence of the MCMC chains.

We use 3 sampling chains with zero divergance.

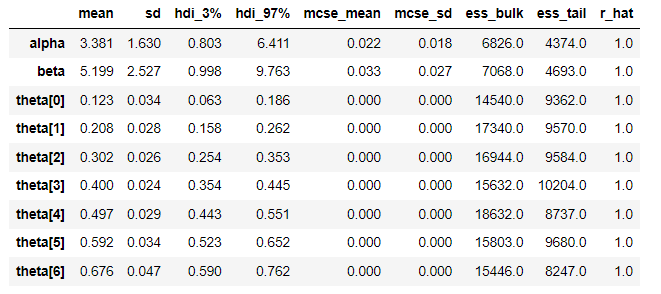


**Question 4 c:**

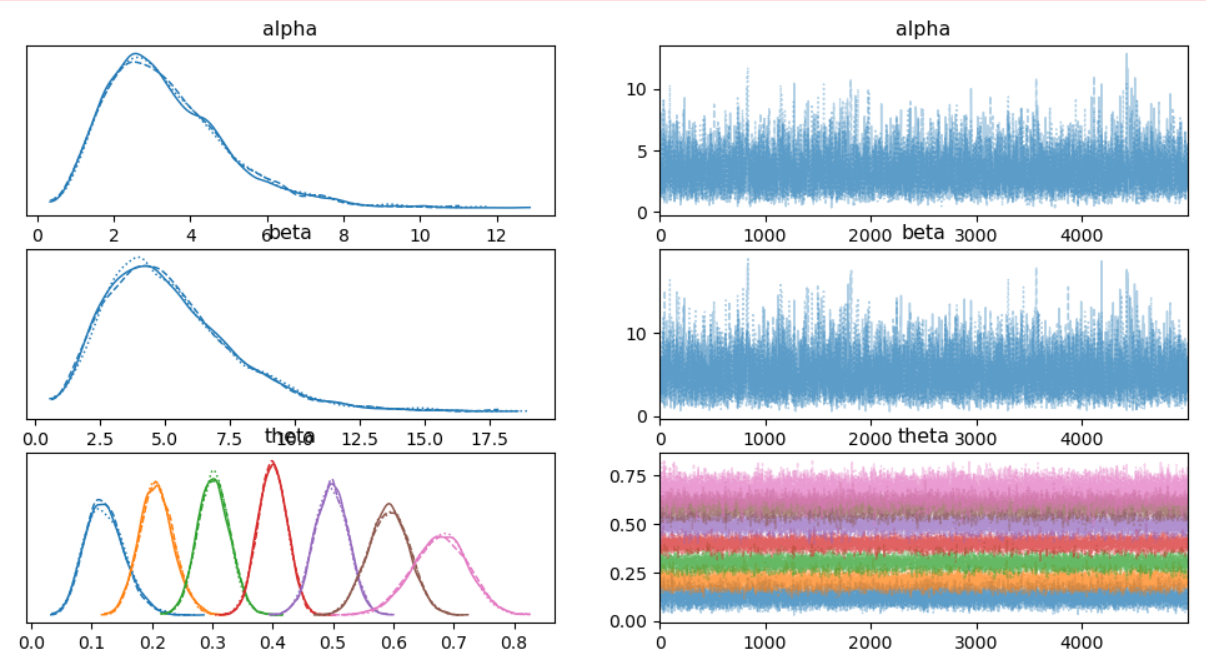
Here, the columns can be represented as:

1. mean: the mean of the posterior distribution of the parameter
2. sd: the standard deviation of the posterior distribution of the parameter
3. hdi\_3%: the lower bound of the 3% highest density interval (HDI) of the posterior distribution of the parameter
4. hdi\_97%: the upper bound of the 97% highest density interval (HDI) of the posterior distribution of the parameter
5. mcse\_mean: the estimated standard error of the mean of the parameter
6. mcse\_sd: the estimated standard error of the standard deviation of the parameter
7. ess\_bulk: the estimated effective sample size of the parameter, taking into account the autocorrelation between samples
8. ess\_tail: the estimated effective sample size of the parameter, taking into account the tail of the posterior distribution
9. r\_hat: the Gelman-Rubin statistic, which measures the convergence of the chains to the same target distribution.

The inference obtained from Model 1 is:



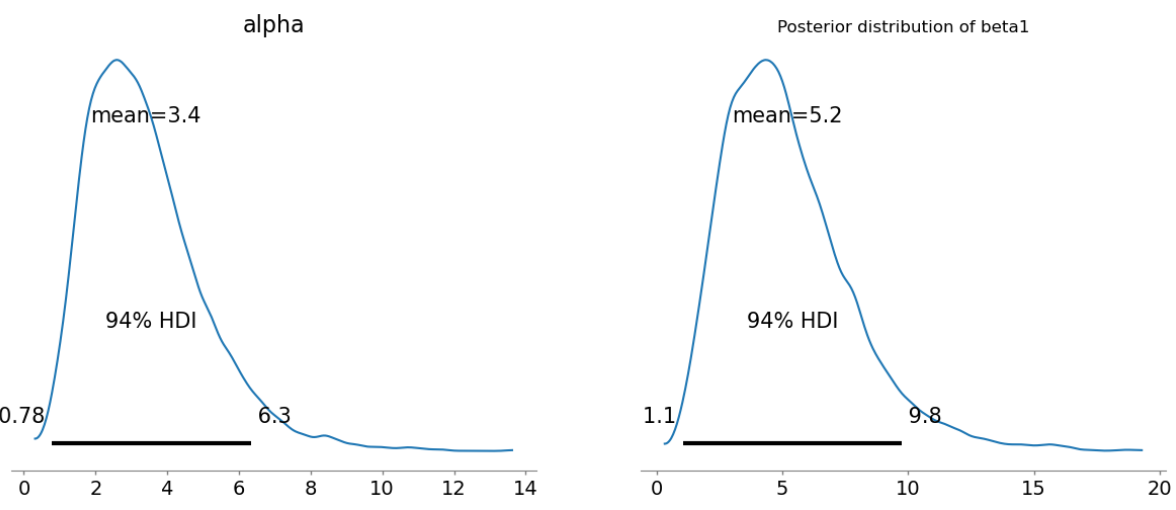
From the above model we can infer that



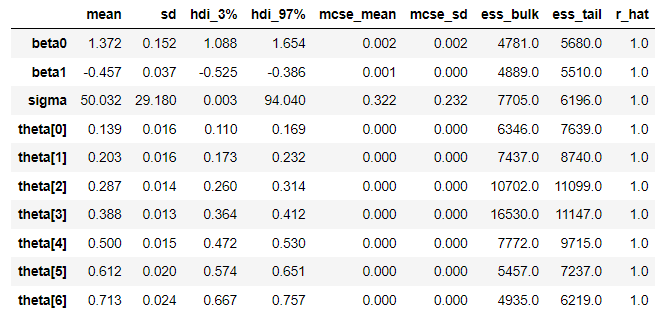
1. Both the sampled of aplha and beta have postive values

2. District 7 i.e thteta[6] has the highest mean (probability of yes) and highest satndard deviation

3. Both Alpha and Beta are left skewed.



The inference for Model 2 is:



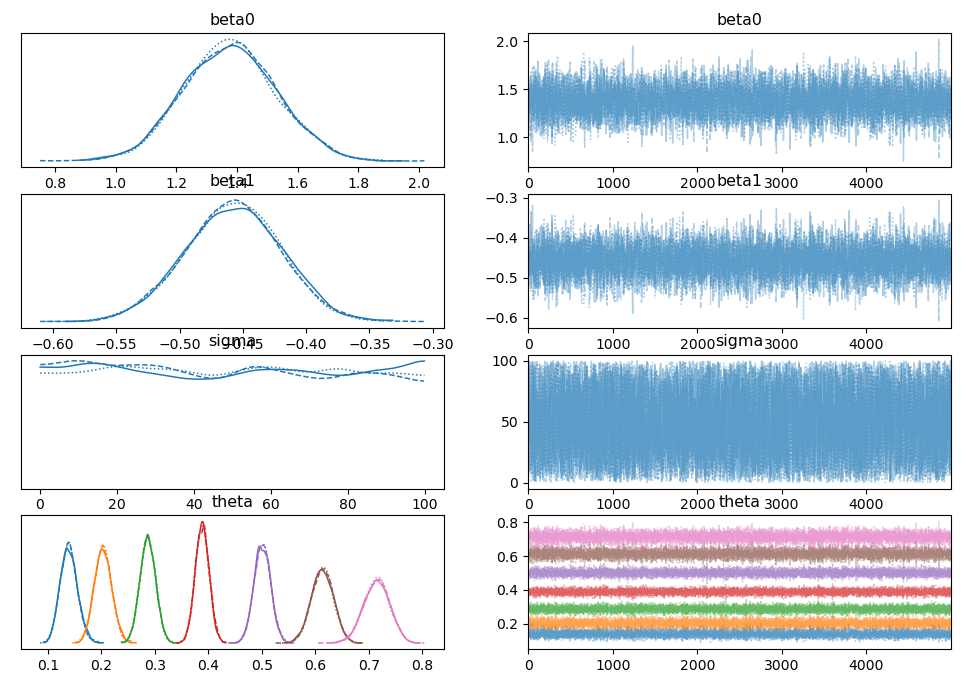
From the above model we can infer that:

1. The samples in beta0 have a positive mean.

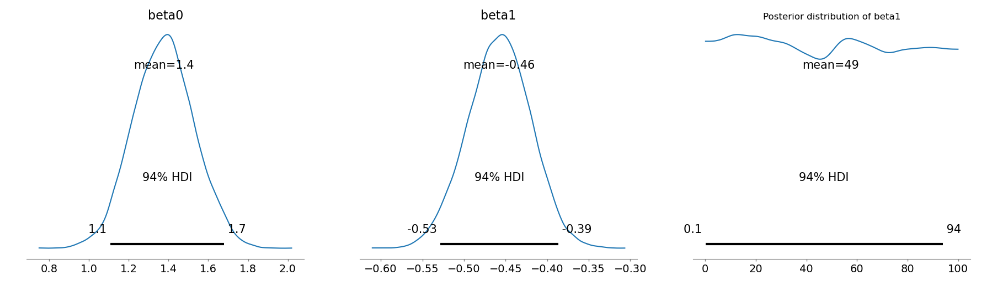
2. The samples in beta1 have a negative mean.

3. District 7 i.e theta[6] has the highest mean and standard deviation which means they have the most number of positve (yes) responses.

The distribution can be given as:



The curve of the normal distribution is bell-shaped, which means that the values in the center of the distribution are more likely to occur than the values at the extremes. The plot of the posterior distribution of beta0 and beta1 looks like a bell-shaped curve it this is because the true values of beta0 and beta1 are close to the center of the distribution.



**Question 4 d:**

In Model 2, β1 represents the effect of the distance between a district and the probability of a ‘yes’ response. Specifically, β1 measures the change in the log-odds of a positive response for unit increase in distance.

If β1 is positive, then the log-odds of a positive response increase with distance. The probability of a positive response decreases as the distance from the clinic increases. Conversely, if β1 is negative, then the log-odds of a positive response decrease with distance, and the probability of a positive response increases as the distance from the clinic increases.

The estimated value of beta1 is -0.457 with a standard deviation of 0.037. This means that in the logistic regression model, for every one unit increase in the predictor variable d, the log-odds of a "yes" response decrease by an estimated 0.457 units on average. Alternatively, we could say that a one-unit increase in d is associated with a decrease in the odds of a "yes" response by a factor of exp(-0.457) = 0.633 on average.

Therefore, β1 captures the relationship between distance and the response variable, and its sign can indicate whether the clinic's location is a significant factor in determining the probability of a positive response.