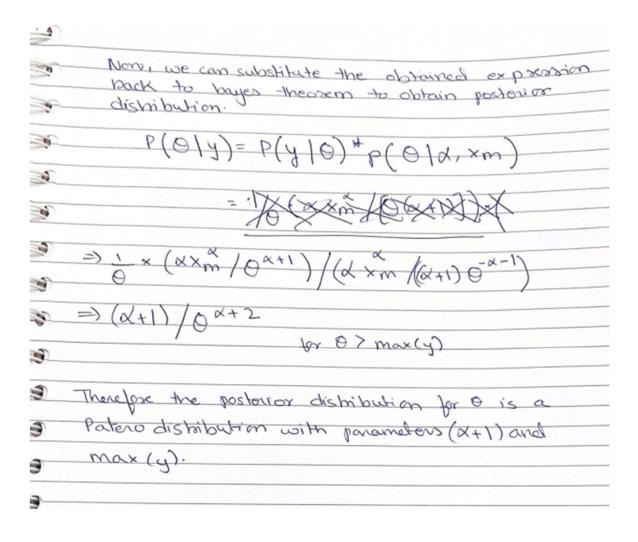
# **Statistics Assignment 3**

Name: Danita Anubhuti Prakash

studentID: **s4745511** 

# Question 1:

| Griven:  Gri   |             |   |
|--|-------------|---|
| Gaiven:  Conform Rondom Variables Vi where i=1, n  and pidif for each as:  P(y10) = 1/0  Pateno Dishibution:  P(x/d,xm) = \( \frac{1}{2} \times \times \)  P(x/d,xm) = \( \frac{1}{2} \times \times \times \)  P(x/m) = \( \frac{1}{2} \times \)  P(y)  | gust        | ion! : Conjugate Uniform Rouxborn Variable Dralysis.  |
| Palero Dishibution:  P(y)0) = 1/0  Palero Dishibution:  P(x d,xm) = & xxm  P(y) = P(y d) * P(0) / P(y)  Where P(y) is the marginal likelihood of total  and it can be given as  P(y) = P(y d) * P(d)dd  Dece can simplify this by substituting the patero dishibution, i.e.  P(y) = P(y d) * P(d xxm)dd  P(y) = P(y d) * P(d xxm)dd  = Xxm  P(y) = & Axm  P(   |             |   |
| P(y) = 1/0  Padero Dishibution:  P(x/d,xm) = \( \text{d} \text{xm} \)  P(x/d) = \( \text{posterior} \)  P(x/d) = \( \text{posterior} \)  P(y) = \( \text{posterior} \)   |             |   |
| P(y) = 1/0  Padero Dishibution:  P(x/d,xm) = { dxm } n > x < xm  P(x/d,xm) = { dxm } n > x < xm  To prove: The posterior dishibution of 0.  Proof:  From the Bayes Theorem, we know that P(0/y) = P(y+0) * P(0) / P(y)  where P(y) is the marginal likelihood of data.  and it can be given as  P(y) = {P(y 0) * P(0)d0}  we can simplify this by substituting the patero dishibution, i.e.  P(y) = \$P(y 0) * P(0/xxm)d0  = \$\frac{xm}{2} \frac{x}{2} \frac{x}{2} \frac{x}{2} \frac{x}{4} \frac{x}{2} \frac{x}{2} \frac{x}{4} \frac{x}{2} \frac{x}{4}  | (           | niform Roundon Variables ti where i=1, n  |
| Padero Dishibution:  P(x/d,xm) = \{ \alpha x \alpha m}  P(x/d) = \{ \alpha (y/d) * \{ \alpha (0) \} \{ \alpha (y)}  where P(y) is the manginal likelihood of data.  and it can be given as  P(y) = \{ \alpha (y/d) * \{ \alpha (0) \} \{ \alpha (0) \}  P(y) = \{ \alpha (y/d) * \{ \alpha x \alpha \} \{ \alpha x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \{ \alpha x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (0/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (y/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (y/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (y/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (y/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (y/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (y/d) x \alpha \} \}  P(y) = \{ \alpha (y/d) * \{ \alpha (y/d) x \alpha \} \}  P(y) = \{ \   |             | and b.g. far each as;   |
| P(x/d,xm)= & x < xm  To prove: The posterior distribution of 0  Proof:  From the Bayes Theorem, we know that P(Oly)=P(y10)* P(0) /P(y)  where P(y) is the marginal likelihood of data.  and it can be given as  P(y)= P(y10) * P(0) d0  we can simplify this by substituting the pateno distribution, i.e.  P(y)= SP(y10) * P(Olaxm) d0  = Xxm O xm / Qx+1 d0  |             | P(y)@)=1/0  |
| P(x/d,xm)= \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \   |             | Palero Dishibution:   |
| Front the Bayes Theoxon, we know that  P(Oly)=P(910)* P(0) /P(y)  where P(y) is the manginal likelihood of data.  and it can be given as  P(y)= P(y10) * P(0)d0  we can simplify this by substituting the patero dishibution, i.e.  P(y)= P(y10) * P(0 xxm)d0  = Xxm P(y10) * P(0 xxm)d0  = Xxm P(y10) * P(0 xxm)d0  |             | C d a n //x   |
| From the bayes Theoxon, we know that  P(O(y) = P(y)O) * P(O) / P(y)  where P(y) is the manginal likelihood of data.  and it can be given as  P(y) = P(y)O) * P(O)dO  we can simplify this by substituting the patero distribution, i.e.  P(y) = P(y)O) * P(O(xxm)dO  = Xxm  O Xm  O  | 0 85 /5/2   | P(x/d,xm)= 3 0 12 m   |
| From the bayes Theorem, we know that  P(O(y) = P(y 10)* P(0) / P(y)  where P(y) is the manginal likelihood of data.  and it can be given as  P(y) = P(y 10) * P(0) d0  we can simplify this by substituting the patero dishribution, i.e.  P(y) = P(y 10) * P(0   axm) d0  = Sm   *(xxm) d0  = Xm   O d0  = Xxm   O d0   |             | ( O x < xm  |
| From the bayes Theorem, we know that  P(O(y) = P(y 10)* P(0) / P(y)  where P(y) is the manginal likelihood of data.  and it can be given as  P(y) = P(y 10) * P(0) d0  we can simplify this by substituting the patero dishribution, i.e.  P(y) = P(y 10) * P(0   axm) d0  = Sm   *(xxm) d0  = Xm   O d0  = Xxm   O d0   | dismission, | DAY OF MELL PROLES  |
| Exom the Bayes Theorem, we know that  P(O(y) = P(y 10)* P(0) / P(y)  where P(y) is the manginal likelihood of data.  and it can be given as  P(y) = P(y 10) * P(0) d0  we can simplify this by substituting the patero dishribution, i-e  P(y) = P(y 10) * P(0   axm) d0  = Sm   *(xxm) d0  = Xm   O d-2 d0  | To p        | move: The posterior distribution of Q.  |
| From the Bayer Theorem, we know that  P(Oly) = P(y 10) * P(0) / P(y)  where P(y) is the manginal likelihood of data.  and it can be given as  P(y) = P(y 10) * P(0) d0  we can simplify this by substituting the patero dishibution, i.e  P(y) = P(y 10) * P(O[xxm] d0  = Xxm P(y) = Xxm A Ox+1 d0  = Xxm A Ox+1 d0  |             | 2 Zerost 1 - Branco Al Cooker City - Senaturi ya Centra ya Kanana ka Kanana |
| P(O(y)=P(y10)*P(0)/P(y)  where $P(y)$ is the manginal likelihood of dota.  and it can be given as $P(y) = \int P(y10) * P(0) d0$ we can simplify this by substituting the patero distribution, i.e. $P(y) = \int P(y10) * P(0(xxm) d0$ $= \int_{-\infty}^{\infty} \frac{1}{10} * (xxm^2/9^{\alpha+1}) d0$ $= xxm \int_{-\infty}^{\infty} \frac{1}{100} * (xxm^2/9^{\alpha+1}) d0$ $= xxm \int_{-\infty}^{\infty} \frac{1}{100} * (xxm^2/9^{\alpha+1}) d0$  | two 1       | Le production of AV down at 1313  |
| P(O(y)=P(y(0)*P(0))P(y)  where $P(y)$ is the manginal likelihood of dota.  and it can be given as $P(y) = \int P(y(0)) * P(0) d0$ we can simplify this by substituting the patero distribution, i.e. $P(y) = \int P(y(0)) * P(0(xxm)) d0$ $= \int_{\infty}^{\infty} \int_{\infty}^{\infty} (x \times m^{\alpha} / 0^{\alpha+1}) d0$ $= \times x \times m \int_{\infty}^{\infty} 0 \times x = 2 d0$   | Comme       | Ata 0 71  |
| where $P(y)$ is the manginal likelihood of data.  and it can be given as $P(y) = \int P(y \mid 0) * P(0) d0$ we can simplify this by substituting the patero distribution, i-e $P(y) = \int P(y \mid 0) * P(0 \mid x \mid x \mid d0)$ $= \int \int \int (x \mid x \mid x \mid x \mid d0)$ $= X \times m \int \int (x \mid x \mid x \mid d0)$ $= X \times m \int \int (x \mid x \mid x \mid d0)$  | 1 80111     | The bayes theosem, we know that   |
| P(y) = $\int P(y 0) * P(0)d0$ we can simplify this by substituting the patero distribution, i.e.  P(y) = $\int P(y 0) * P(0 xxm)d0$ = $\int \frac{1}{2} * (x x m^{2} / 0^{x+1})d0$ = $x \times m$ = $x \times m$   |             | b(0/2)=b(d/0)* b(0) \b(d)   |
| P(y) = $\int P(y 0) * P(0)d0$ we can simplify this by substituting the patero distribution, i.e.  P(y) = $\int P(y 0) * P(0 xxm)d0$ = $\int \frac{1}{2} * (x \times m^{2} / 0^{2} + 1) d0$ = $x \times m$ = $x \times m$   | when        | P(y) is the marginal likelihood or data.  |
| $P(y) = \int P(y 0) * P(0)d0$ we can simplify this by substituting the patero distribution, i-e $P(y) = \int P(y 0) * P(0 x \times m) d0$ $= \int_{-\infty}^{\infty} \frac{1}{m} * (x \times m) * (x \times$ | and         | it can be given as  |
| be can simplify this by substituting the patero distribution, i-e  P(y) = \[ P(y 0) \times p(0 xxm) d0 \]  = \[ \frac{xm}{2} \frac{1}{2} \frac{x}{2} \frac{1}{2} \   | 1.70        |   |
| we can simplify this by substituting the patero distribution, i.e. $P(y) = \int P(y \theta) p(\theta xxm) d\theta$ $= \int_{xm}^{xm} (xxm^{\alpha}/\theta^{\alpha+1}) d\theta$ $= x \times m \int_{xm}^{xm} (x^{\alpha}/\theta^{\alpha+1}) d\theta$  |             | P(y)= \P(y)0) *P(0)40   |
| $P(y) = \int P(y 0) * P(0 x \times m) d0$ $= \int \frac{1}{2\pi} * (x \times m^{\alpha} / 0^{\alpha+1}) d0$ $= x \times m \int 0 x^{\alpha-2} d0$  |             | Ö   |
| $P(y) = \int P(y 0) * P(0 x \times m) d0$ $= \int \frac{1}{2\pi} * (x \times m^{\alpha} / 0^{\alpha+1}) d0$ $= x \times m \int 0 x^{\alpha-2} d0$  | we c        | an simplify this by substituting the  |
| $P(y) = \int P(y 0) * P(0 x \times m) d0$ $= \int \frac{1}{2\pi} * (x \times m^{\alpha} / 0^{\alpha+1}) d0$ $= x \times m \int 0 x^{\alpha-2} d0$  | pater       | o distribution, i-e designa of the book   |
| $= \frac{1}{2} \frac{1}{8} \frac{1}{4} \left( \frac{1}{2} \times \frac{1}{2} \right) d\theta$ $= \frac{1}{2} \times \frac{1}{4} \left( \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \right) d\theta$  |             | 1 2 Comment of the Allen and Secretary  |
| $= \frac{1}{2} \frac{1}{8} \frac{1}{4} \left( \frac{1}{2} \times \frac{1}{2} \right) d\theta$ $= \frac{1}{2} \times \frac{1}{4} \left( \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \right) d\theta$  | P           | (y)=) P(y10) P(0 xxm) d0  |
| = xxm (0 d0  |             | O Xm  |
| = xxm (0 d0  |             | = ) = (xxm / (x+1) do   |
| = xxm \ 0 \ d0   | 1. 1. 100   | xm - 2 - 2  |
| - X X / X - X - 1]   | - W         | = xxm (0 d0   |
| = X X / X 1 1 10 - X - 1   |             | , O   |
|  |             |   |



## Question 2:

| Juestion 2:  |
|--|
| Griven mild alegorical variables, Yi, (i=1,-m) each with p.d. f P(y) P1 px) = The The prior for O= (p1 px) is the divichet dibibution (xor, Xo where Ko is the hyper-parameter for the jth category.   |
| Given nied antegorical variables. Yi (1:1, -in) each   |
| with p.d. P (y) P1 PX) = II o.   |
| William in to la la land   |
| The prior for O: (p1 bK) is the divichet dividuos  |
| dot, do where do is the hyper-parameter for the  |
| it category.   |
|  |
| A BOD 310 B Problem only sill of a proper  |
| 10 prox stocked Derive the posterior distribution of O   |
| Solution is solve on a control of the constron and   |
| Solution: 2 200 1000 1000 100 100 100  |
| The likelihood didinibility in it the first with the of  |
| the o observations ushich is the point distribution of   |
| probabilities for perfected as it is   |
| K and the second |
| million application of the principle windson will  |
| and Man Playant borlobder others   |
| The likelihood dishribation is the joint dishribation of the n observations which is the product of individual probabilities of the product of individual  R  (Of grade yn) = TT (Pi)  |
| where ny is the number of observations in category i.  |
|  |
| the prior distribution is:   |
| The prior distribution is:  P(O) = Direct ( 201 do)  |
| The postprior distribution is the mountained to the  |
| The posterior distribution is the proportional to the product of the likelihood function and the prior distribution. That is   |
| distribution. That is:   |
|  |
| P(0) y1, yn) x L(0) y1 -, yn) * p(0)   |
|  |
| Taking log on both sides, we get   |
| log ((p) g1 yn)) × log(1(0) y1, yn))+ log φ(0))  |
|  |
|  |
|  |

| · C midaan  | 1      |
|---|--------|
| Using the proporties of the logarithm, we can simple this to:  log(p(0 y1yn)) X  kg  Kart  Ka | 13     |
| their of the properties of the  | (e)    |
| MS 10: 11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   | 3F - 1 |
| 195   |        |
| log(p(0)y,,yn))X  | 7-     |
| dulch town with Kart or Cart 19 100   |        |
| all of mex local and a factor   | ,      |
| J=1 (n; x (og(p;))) + & (a6-1)  |        |
| J=1   |        |
| i i - 1 1:1:14 of category  |        |
| where p; is the probability of the word of  |        |
| a continuos be minorizare est or dischlet dishibi   | wion   |
| where p; is the probability of category is a directlet distribution is a directlet distribution with applicated hyperparameters   |        |
| with applated hyperparameters   |        |
| 11. (2) (1) (1)   |        |
| ancol ( un) = Dirichlet ( dy - 1 dx)  |        |
| of the state of it did and and and and and and and and and an   |        |
| where p: is the probability of category judge   |        |
| where prosent   |        |
| The posterior distribution is a dirichlet distribution with updated hyperparameters   | m      |
| The posterior distribution  |        |
| with updated hyperpararres  |        |
| 10- (du +-10d 14) 100   |        |
| o(O) y - yn) = Dinchler Cont  |        |
| where $\alpha_j = \alpha_0^2 + \alpha_j$  |        |
| N: = Xi + Di mal mortadassas  | 200/   |
| where a discolor  | -      |
| lavier distribution of O is a district  |        |
| so the posterior wholated human parameters the  |        |
| distribution with appearant of the so but one   |        |
| in - conorate the observed  |        |
| so the posterior distribution of O is a dissolved distribution with updated hyper-parameters that in-corporate the observed date.   |        |
| 10/0x(0B-18/9) 1 × (0P - NP/9) 4  |        |
| - L & ( , W , . , W   O ) - 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   |        |

## Question 3 a:

| Question 3   | 7         |
|--|-----------|
| Suestion 3   |           |
| Criven:  |           |
| f(x,y) = ce (xy+x+y) x>0, y>0  |           |
| for a normalization constant c, using gibbs                                      |           |
| Sampler. Such that (x, y) ~ }  |           |
| to prove:  |           |
| The anditional probability of x given Y.y. and the andition toll of x given X=X. | 6         |
|  |           |
| Solution:  |           |
|  | 6-2       |
| Part - 1: To find the conditional pdf of X given                                 | 7=4       |
| Part - 1: To find the conditional palf of X given we use Bayes Theorem.          |           |
| f(x/d)=f(x,d)/f(d)   | 3/61      |
| Part - 2 - To lind the conditional poly of y give                                | n X = X   |
| Part - 2 = To find the conditional pall of Y give                                | 22        |
| P(xx1x) = P(x, x) P(x)   | l life    |
| First we lind the joint distribution of 1(x, y)                                  |           |
| First we find the joint distribution of f(x,y)                                   |           |
|  | Carena di |
| Now we can find the marginal distribution of X.                                  | - 4       |
| P(A): If(x, A) 4x = 2 co-(xA+x+A) 4x   |           |
| = (c=18-x-9 dx = ce-9 [e-(x+1) 9 d   | ×         |
|  |           |
| = ce-y[-1/y+1].e-cx+1)y+c  |           |
| = ce-4/1+4   |           |
|  |           |

| Next, we find the conditional distribution of x given Y=y         |
|---|
| f(x/y)= f(x, y) /f(y)   |
| = ce-xy-x-3/(ce-y/1+y)  |
| = (0000 e* (x(y+1))/(1+y)   |
| = [e (x(g+11))] (1+y)   |
| Similarly, the coorditional distribution of y given X=X           |
| P(A1x)= f(x, A) / f(x)  |
| = ce-(xyxx+y) /(ce-x/1+x)   |
| = (XXXXX) e-y(x+1)/(e-x/1+x)                                      |
| $= e^{-y(x+1)}/(1+x)$   |
| In summary, the conditional poly of x given Yey is ex (y+1) (1+9) |
| and the conditional poly of I given x=x i's                       |
| e-A(x+1) (1+x)  |
|   |
|   |

#### 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  $(X, Y) \sim f$ . The working code that implements the Gibbs sampler and outputs 1000 points that are approximately distributed according to f is as follows:

```
import numpy as np
  # Define the Gibbs sampler function
  def gibbs sampler(num samples):
      x, y = 1, 1 # Starting values of x and y
      samples = np.zeros((num_samples, 2)) # Initialize array to store samples
      for i in range(num samples):
          # Sample x from the conditional distribution p(x|y)
          x = np.random.gamma(2, 1/(y+1))
          # Sample y from the conditional distribution p(y|x)
          y = np.random.gamma(2, 1/(x+1))
          samples[i, :] = [x, y] # Store the sample
      return samples
  # Run the Gibbs sampler to generate 1000 samples
  samples = gibbs_sampler(1000)
  # Print the first 10 samples
  print(samples[:10, :])
```

#### **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.

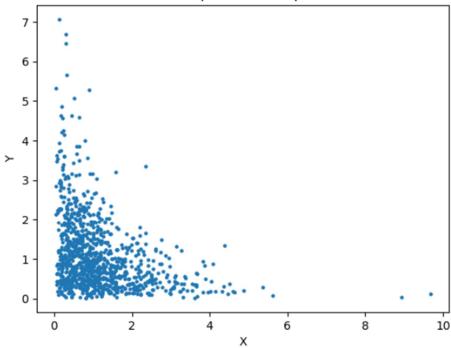
In each iteraton, a new value of x and y is sampled from the function np.random.gamma(2, 1/(y+1)) and np.random.gamma(2, 1/(x+1)). This is then stored in the samples array and returned. Here we do not use the f(x, y) function, but it is implicitly used by the gibbs distibution.

#### **OUTPUT:**

```
[[1.28047846 0.74544487]
[0.88104641 0.70974176]
[2.13390308 0.30993926]
[4.50020334 0.37779603]
[1.66773187 0.57691352]
[2.49341698 0.56281007]
[3.14283084 1.30936054]
[0.46349223 0.70674931]
[4.3887477 1.33856779]
[0.55011417 3.12884922]]
```

```
# Plot the sampled points
plt.scatter(samples[:, 0], samples[:, 1], s=5)
plt.xlabel("X")
plt.ylabel("Y")
plt.title("Samples from 2D pdf")
plt.show()
```

### Samples from 2D pdf



### **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:

```
⋈ import pymc3 as pm

  import numpy as np
  # Define the data
  r = np.array([10, 40, 90, 160, 150, 120, 70]) # number of YES answers in each district
  n = np.array([100, 200, 300, 400, 300, 200, 100]) # total number of responses in each district
  J = len(r) # number of districts
  # Define the model
  with pm.Model() as model:
      # Priors for hyperparameters
      alpha = pm.Uniform('alpha', lower=0, upper=100)
      beta = pm.Uniform('beta', lower=0, upper=100)
      # Parameters of interest
      theta = pm.Beta('theta', alpha=alpha, beta=beta, shape=J)
      # Likelihood
      r_obs = pm.Binomial('r_obs', n=n, p=theta, observed=r)
      # MCMC settings
      trace model1 = pm.sample(draws=5000, tune=1000, chains=3)
  # Print the summary of the posterior distributions
  pm.summary(trace_model1)
```

#### **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:

```
import numpy as np
  import pymc3 as pm
  # Define the data
  n = np.array([100, 200, 300, 400, 300, 200, 100])
  r = np.array([10, 40, 90, 160, 150, 120, 70])
  d = np.array([7, 6, 5, 4, 3, 2, 1])
  # Define the model
  with pm.Model() as model:
      # Priors for the parameters
      beta0 = pm.Uniform('beta0', lower=-10, upper=10)
      beta1 = pm.Uniform('beta1', lower=-10, upper=10)
      sigma = pm.Uniform('sigma', lower=0, upper=100)
      # Prior of theta
      theta = pm.math.invlogit(beta0 + beta1 * d)
      # #The likelihood of the number of YES answers in group j, r_j
      likelihood = pm.Binomial('likelihood', n=n, p=theta, observed=r)
      # Define the theta variable and add it to the trace.
      theta_var = pm.Deterministic('theta', theta)
      # Sample from the posterior distribution
      trace_model2 = pm.sample(5000, chains=3, target_accept=0.9)
  pm.summary(trace_model2)
```

#### **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.



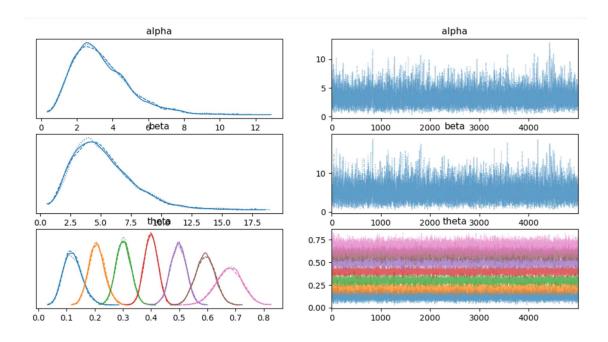
#### Question 4 c:

Here, the columns can be represented as:

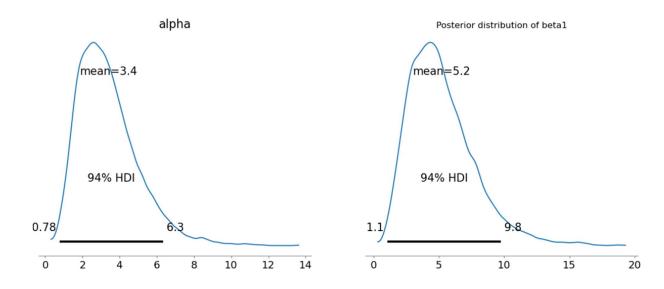
- 1. mean: the mean of the posterior distribution of the parameter
- 2. sd: the standard deviation of the posterior distribution of the parameter
- 3. hdi\_3%: the lower bound of the 3% highest density interval (HDI) of the posterior distribution of the parameter
- 4. hdi\_97%: the upper bound of the 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:

|          | mean  | sd    | hdi_3% | hdi_97% | mcse_mean | mcse_sd | ess_bulk | ess_tail | r_hat |
|----------|-------|-------|--------|---------|-----------|---------|----------|----------|-------|
| alpha    | 3.381 | 1.630 | 0.803  | 6.411   | 0.022     | 0.018   | 6826.0   | 4374.0   | 1.0   |
| beta     | 5.199 | 2.527 | 0.998  | 9.763   | 0.033     | 0.027   | 7068.0   | 4693.0   | 1.0   |
| theta[0] | 0.123 | 0.034 | 0.063  | 0.186   | 0.000     | 0.000   | 14540.0  | 9362.0   | 1.0   |
| theta[1] | 0.208 | 0.028 | 0.158  | 0.262   | 0.000     | 0.000   | 17340.0  | 9570.0   | 1.0   |
| theta[2] | 0.302 | 0.026 | 0.254  | 0.353   | 0.000     | 0.000   | 16944.0  | 9584.0   | 1.0   |
| theta[3] | 0.400 | 0.024 | 0.354  | 0.445   | 0.000     | 0.000   | 15632.0  | 10204.0  | 1.0   |
| theta[4] | 0.497 | 0.029 | 0.443  | 0.551   | 0.000     | 0.000   | 18632.0  | 8737.0   | 1.0   |
| theta[5] | 0.592 | 0.034 | 0.523  | 0.652   | 0.000     | 0.000   | 15803.0  | 9680.0   | 1.0   |
| theta[6] | 0.676 | 0.047 | 0.590  | 0.762   | 0.000     | 0.000   | 15446.0  | 8247.0   | 1.0   |



- 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.



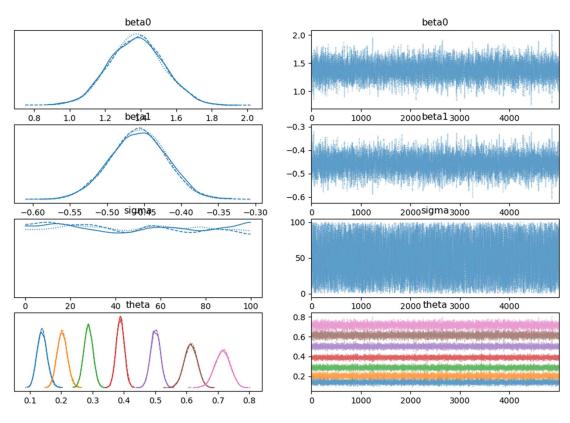
#### The inference for Model 2 is:

| _        | mean   | sd     | hdi_3% | hdi_97% | mcse_mean | mcse_sd | ess_bulk | ess_tail | r_hat |
|----------|--------|--------|--------|---------|-----------|---------|----------|----------|-------|
| beta0    | 1.372  | 0.152  | 1.088  | 1.654   | 0.002     | 0.002   | 4781.0   | 5680.0   | 1.0   |
| beta1    | -0.457 | 0.037  | -0.525 | -0.386  | 0.001     | 0.000   | 4889.0   | 5510.0   | 1.0   |
| sigma    | 50.032 | 29.180 | 0.003  | 94.040  | 0.322     | 0.232   | 7705.0   | 6196.0   | 1.0   |
| theta[0] | 0.139  | 0.016  | 0.110  | 0.169   | 0.000     | 0.000   | 6346.0   | 7639.0   | 1.0   |
| theta[1] | 0.203  | 0.016  | 0.173  | 0.232   | 0.000     | 0.000   | 7437.0   | 8740.0   | 1.0   |
| theta[2] | 0.287  | 0.014  | 0.260  | 0.314   | 0.000     | 0.000   | 10702.0  | 11099.0  | 1.0   |
| theta[3] | 0.388  | 0.013  | 0.364  | 0.412   | 0.000     | 0.000   | 16530.0  | 11147.0  | 1.0   |
| theta[4] | 0.500  | 0.015  | 0.472  | 0.530   | 0.000     | 0.000   | 7772.0   | 9715.0   | 1.0   |
| theta[5] | 0.612  | 0.020  | 0.574  | 0.651   | 0.000     | 0.000   | 5457.0   | 7237.0   | 1.0   |
| theta[6] | 0.713  | 0.024  | 0.667  | 0.757   | 0.000     | 0.000   | 4935.0   | 6219.0   | 1.0   |

From the above model we can infer that:

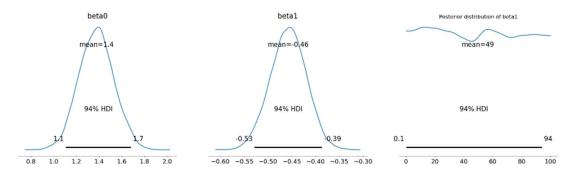
- 1. The samples in beta0 have a positive mean.
- 2. The samples in beta1 have a negative mean.
- 3. District 7 i.e theta[6] has the highest mean and standard deviation which means they have the most number of positve (yes) responses.

### 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,  $\beta$ 1 represents the effect of the distance between a district and the probability of a 'yes' response. Specifically,  $\beta$ 1 measures the change in the log-odds of a positive response for unit increase in distance.

If  $\beta 1$  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  $\beta 1$  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 logodds 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,  $\beta 1$  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.