BDA Assignment

For one of the problems and models described below, you are required to implement python programs in **Pymc3** using **MCMC and variational inference** algorithms to learn posterior distribution over the parameters (including hyperparameters and other unknown parameters). You need to provide samples from the posterior and various statistics like mean, median and variance. You are also required to compare your results with various MCMC and VI routines and compare the results. The learnt posterior distribution is then used to make predictions on either a held out test data points or training data points (if the data is very small) and report the error. Please submit both code and a report discussing the details (proposal distribution, approximate distribution, initialization, burning in period, mathematical expression for the lower bound used by VI in PyMC3 for your problem and model, details of the samplers used in PyMC3 and sampling process, parameters of sampler, posterior distribution plots), experimental setup, observations and results for the assignment. Data for the following problems can be found here : https://drive.google.com/drive/folders/1cGTzfKrIOU_8K4v_2gRVuuMEerSpl-YO?usp=sharing

Each problem can be picked by a maximum of 3 groups only and each group can have max 3 members. Please form your group and fill in the following form with your group member details and one chosen assignment question. Assignment problems will be allocated on first come first serve basis (once 3 groups choose a problem, please choose another problem)
https://docs.google.com/spreadsheets/d/1hXBDOGOXQCGVCxMClCgtbyaJqQy8-0Zm-hqNB3w7wxk/edit?usp=sharing

Problem - 1 : Dugong

Carlin and Gelfand (1991) present a nonconjugate Bayesian analysis of the following data set from Ratkowsky (1983):

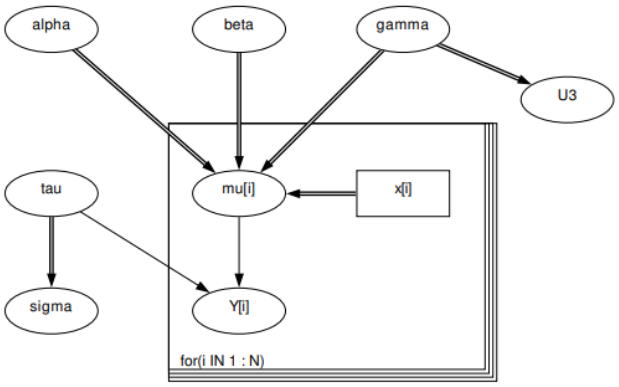
Dugong	1	2	3	4	5	....	26	27
Age (X)	1.0	1.5	1.5	1.5	2.5	....	29.0	31.5
Length (Y)	1.80	1.85	1.87	1.77	2.02	....	2.27	2.57

The data are length and age measurements for 27 captured dugongs (sea cows). Carlin and Gelfand (1991) model this data using a nonlinear growth curve with no inflection point and an asymptote as X_i tends to infinity:

$$Y_i \sim \text{Normal}(\mu_i, \tau), \quad i = 1,...,27$$

$$\mu_i = \alpha - \beta\gamma^{X_i} \quad \alpha, \beta > 1; 0 < \gamma < 1$$

Standard noninformative priors are adopted for α, β and τ, and a uniform prior on (0,1) is assumed for γ. However, this specification leads to a non conjugate full conditional distribution for γ which is also non log-concave. The graph and corresponding BUGS code is given below




```
model
{
   for( i in 1 : N ) {
      Y[i] ~ dnorm(mu[i], tau)
      mu[i] <- alpha - beta * pow(gamma,x[i])
   }
   alpha ~ dnorm(0.0, 1.0E-6)
   beta ~ dnorm(0.0, 1.0E-6)
   gamma ~ dunif(0.5, 1.0)
   tau ~ dgamma(0.001, 0.001)
   sigma <- 1 / sqrt(tau)
   U3 <- logit(gamma)
}
```

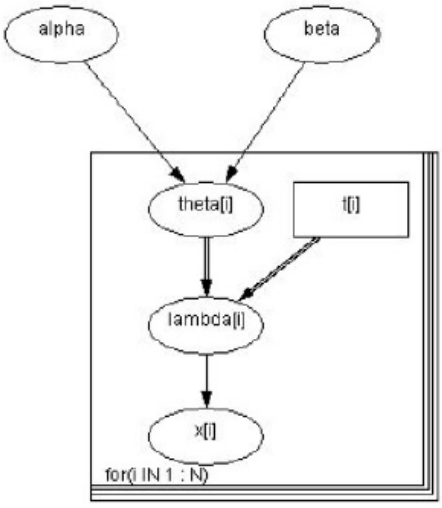
Problem 2 : Pumps

You have 10 power plant pumps. For these 10 pumps, you are given the number of failures (x_i) and the number of hours in thousands that the pumps have run for (t_i).
They seek to model the relationship between x i and t i using a Gamma-Poisson Hierarchical model as done by George et al in their paper Conjugate Likelihood Distributions in 1993.

$$X_i \sim Poisson(\theta_i t_i),\ i = 1, 2, \ldots, 10 \text{ where } \theta_i \text{ is the failure rate}$$

$$\theta_i \sim \Gamma(a, b), \quad i = 1, 2, \ldots, 10, \qquad a \sim Exponential(1) \quad \text{and} \quad b \sim \Gamma(0.1, 1).$$

Pump	1	2	3	4	5	6	7	8	9	10
t_i	94.5	15.7	62.9	126	5.24	31.4	1.05	1.05	2.1	10.5
x_i	5	1	5	14	3	19	1	1	4	22



`x[i]` is Poisson-distributed with mean `lambda[i]` for every `i` = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

`lambda[i]` is a linearly transformed random variable whose parent random variable is `theta[i]`. Mathematically, `lambda[i]` = `theta[i] * t[i]` for every `i` = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.
`theta[i]` is Gamma-distributed with parameters `alpha` and `beta`.
`alpha` is Exponential-distributed with mean `1`.
`beta` is Gamma-distributed with parameters `0.1` and `1`.

Problem 3 - Blockers

Carlin (1992) considers a Bayesian approach to meta-analysis, and includes the following examples of 22 trials of beta-blockers to prevent mortality after myocardial infarction.

Study	**Mortality: deaths / total**	
	Treated	*Control*
1	3/38	3/39
2	7/114	14/116
3	5/69	11/93
4	102/1533	127/1520
.....		
20	32/209	40/218
21	27/391	43/364
22	22/680	39/674

In a random effects meta-analysis we assume the true effect (on a log-odds scale) δ_i in a trial *i* is drawn from some population distribution.Let r^C_i denote number of events in the control group in trial *i*, and r^T_i denote events under active treatment in trial *i*. Our model is:

$$r^C_i \sim \text{Binomial}(p^C_i, n^C_i)$$

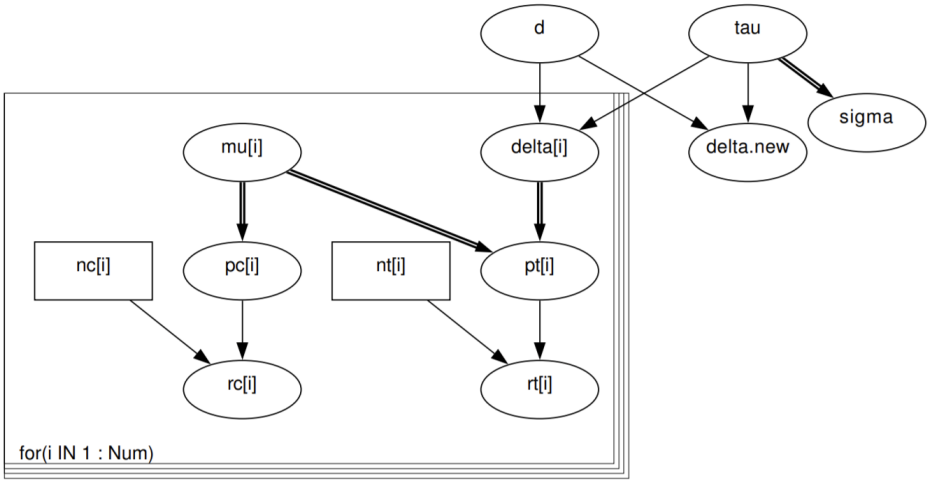
$$r^T_i \sim \text{Binomial}(p^T_i, n^T_i)$$

$$\text{logit}(p^C_i) = \mu_i$$

$$\text{logit}(p^T_i) = \mu_i + \delta_i$$

$$\delta_i \sim \text{Normal}(d, \tau)$$

``Noninformative" priors are given for the μ_i's. τ and d. The graph for this model is shown in




```
model
{
   for( i in 1 : Num ) {
      rc[i] ~ dbin(pc[i], nc[i])
      rt[i] ~ dbin(pt[i], nt[i])
      logit(pc[i]) <- mu[i]
      logit(pt[i]) <- mu[i] + delta[i]
      mu[i] ~ dnorm(0.0,1.0E-5)
      delta[i] ~ dt(d, tau, 4)
   }
   d ~ dnorm(0.0,1.0E-6)
   tau ~ dgamma(0.001,0.001)
   delta.new ~ dt(d, tau, 4)
   sigma <- 1 / sqrt(tau)
}
```

Problem 4 : Beetles

Dobson (1983) analyses binary dose-response data published by Bliss (1935), in which the numbers of beetles killed after 5 hour exposure to carbon disulphide at N = 8 different concentrations are recorded:

Concentration (x_i)	Number of beetles (n_i)	Number killed (r_i)
1.6907	59	6
1.7242	60	13
1.7552	62	18
1.7842	56	28
1.8113	63	52
1.8369	59	52
1.8610	62	61
1.8839	60	60

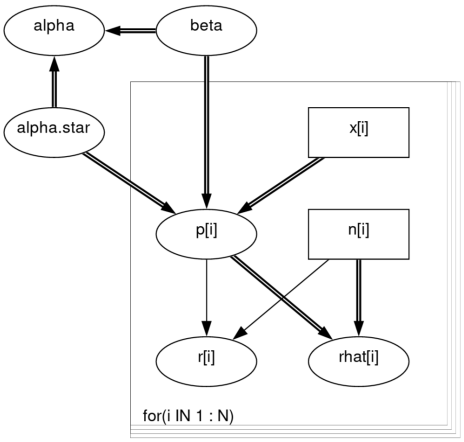
We assume that the observed number of deaths r_i at each concentration x_i is binomial with sample size n_i and true rate p_i. Plausible models for pi include the logistic, probit and extreme value (complimentary log-log) models, as follows

$$p_i = \exp(\alpha + \beta x_i) / (1 + \exp(\alpha + \beta x_i)$$

$$p_i = \mathrm{Phi}(\alpha + \beta x_i)$$

$$p_i = 1 - \exp(-\exp(\alpha + \beta x_i))$$

The corresponding graph is shown below:




```
model
{
   for( i in 1 : N ) {
      r[i] ~ dbin(p[i],n[i])
      logit(p[i]) <- alpha.star + beta * (x[i] - mean(x[]))
      rhat[i] <- n[i] * p[i]
   }
   alpha <- alpha.star - beta * mean(x[])
   beta ~ dnorm(0.0,0.001)
   alpha.star ~ dnorm(0.0,0.001)
}
```

Problem 5 : Dogs

Lindley (19??) analyses data from Kalbfleisch (1985) on the Solomon-Wynne experiment on dogs, whereby they learn to avoid an electric shock. A dog is put in a compartment, the lights are turned out and a barrier is raised, and 10 seconds later an electric shock is applied. The results are recorded as success (Y = 1) if the dog jumps the barrier before the shock occurs, or failure (Y = 0) otherwise.

Thirty dogs were each subjected to 25 such trials. A plausible model is to suppose that a dog learns from previous trials, with the probability of success depending on the number of previous shocks and the number of previous avoidances. Lindley thus uses the following model

$$\pi_j = A^{x_j} B^{j-x_j}$$

for the probability of a shock (failure) at trial j, where x_j = number of success (avoidances) before trial j and $j - x_j$ = number of previous failures (shocks). This is equivalent to the following log linear model

$$\log \pi_j = \alpha x_j + \beta (j-x_j)$$

```
model
{
   for (i in 1 : Dogs) {
      xa[i, 1] <- 0; xs[i, 1] <- 0 p[i, 1] <- 0
      for (j in 2 : Trials) {
         xa[i, j] <- sum(Y[i, 1 : j - 1])
         xs[i, j] <- j - 1 - xa[i, j]
         log(p[i, j]) <- alpha * xa[i, j] + beta * xs[i, j]
         y[i, j] <- 1 - Y[i, j]
         y[i, j] ~ dbern(p[i, j])
      }
   }
   alpha ~ dnorm(0, 0.00001)I(, -0.00001)
   beta ~ dnorm(0, 0.00001)I(, -0.00001)
   A <- exp(alpha)
   B <- exp(beta)
}
```

Problem 6 : Salmonella

Breslow (1984) analyses some mutagenicity assay data (shown below) on salmonella in which three plates have been processed at each dose *i* of quinoline and the number of revertant colonies of TA98 Salmonella measured. A certain dose-response curve is suggested by theory.

dose of quinoline (μg per plate)					
0	10	33	100	333	1000
15	16	16	27	33	20
21	18	26	41	38	27
29	21	33	69	41	42

This is assumed to be a random effects Poisson model allowing for over-dispersion. Let x_i be the dose on the plates *i*1, *i*2 and *i*3. Then we assume

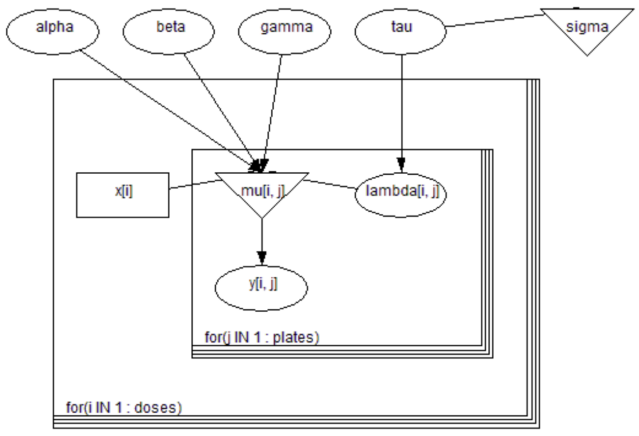
$$y_{ij} \sim \text{Poisson}(\mu_{ij})$$

$$\log(\mu_{ij}) = \alpha + \beta \log(x_i + 10) + \gamma x_i + \lambda_{ij}$$

$$\lambda_{ij} \sim \text{Normal}(0, \tau)$$

α , β , γ , τ are given independent ``noninformative'' priors. The appropriate graph is shown

Graphical model for salm example




```
model
{
  for( i in 1 : doses ) {
    for( j in 1 : plates ) {
      y[i , j] ~ dpois(mu[i , j])
      log(mu[i , j]) <- alpha + beta * log(x[i] + 10) +
        gamma * x[i] / 1000 + lambda[i , j]
      lambda[i , j] ~ dnorm(0.0, tau)
    }
  }
  alpha ~ dnorm(0.0,1.0E-6)
  beta ~ dnorm(0.0,1.0E-6)
  gamma ~ dnorm(0.0,1.0E-6)
  tau ~ dgamma(0.001, 0.001)
  sigma <- 1 / sqrt(tau)
}
```

Problem 7 : Seed

This example is taken from Table 3 of Crowder (1978), and concerns the proportion of seeds that germinated on each of 21 plates arranged according to a 2 by 2 factorial layout by seed and type of root extract. The data are shown below, where r_i and n_i are the number of germinated and the total number of seeds on the *i* th plate, *i* =1,...,N. These data are also analysed by, for example, Breslow: and Clayton (1993).

	seed O. aegyptiaco 75						*seed O. aegyptiaco 73*				
	Bean			Cucumber			Bean			Cucumber	
r	n	r/n	r	n	r/n	r	n	r/n	r	n	r/n
10	39	0.26	5	6	0.83	8	16	0.50	3	12	0.25
23	62	0.37	53	74	0.72	10	30	0.33	22	41	0.54
23	81	0.28	55	72	0.76	8	28	0.29	15	30	0.50
26	51	0.51	32	51	0.63	23	45	0.51	32	51	0.63
17	39	0.44	46	79	0.58	0	4	0.00	3	7	0.43
			10	13	0.77						

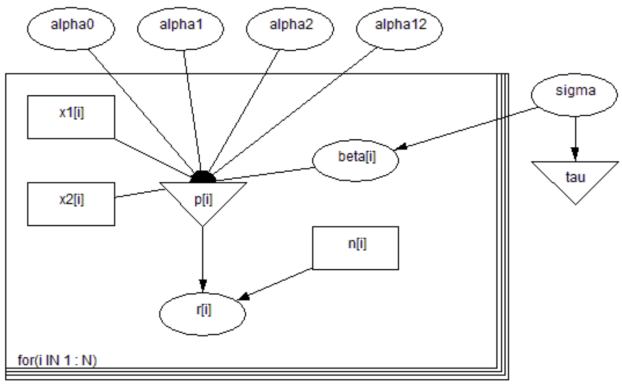
The model is essentially a random effects logistic, allowing for over-dispersion. If p_i is the probability of germination on the *i* th plate, we assume

$r_i \sim \text{Binomial}(p_i, n_i)$

$\text{logit}(p_i) = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_{12} x_{1i} x_{2i} + b_i$

$b_i \sim \text{Normal}(0, \tau)$

where x_{1i} , x_{2i} are the seed type and root extract of the *i* th plate, and an interaction term $\alpha_{12} x_{1i} x_{2i}$ is included. α_0 , α_1 , α_2 , α_{12} , τ are given independent "noninformative" priors.




```
model
{
  for( i in 1 : N ) {
    r[i] ~ dbin(p[i],n[i])
    beta[i] ~ dnorm(0.0,tau)
    logit(p[i]) <- alpha0 + alpha1 * x1[i] + alpha2 * x2[i] +
      alpha12 * x1[i] * x2[i] + beta[i]
  }
  alpha0 ~ dnorm(0.0,1.0E-6)
  alpha1 ~ dnorm(0.0,1.0E-6)
  alpha2 ~ dnorm(0.0,1.0E-6)
  alpha12 ~ dnorm(0.0,1.0E-6)
  sigma ~ dunif(0,10)
  tau <- 1 / pow(sigma, 2)
}
```

Problem 8 : Hearts

The table below presents data given by Berry (1987) on the effect of a drug used to treat patients with frequent premature ventricular contractions (PVCs) of the heart.

number (i)	*PVC's per minute* Pre-drug (x_i)	Post-drug (y_i)	Decrease
1	6	5	1
2	9	2	7
3	17	0	17
.	..	..	..
11	9	13	–4
12	51	0	51

Farewell and Sprott (1988) model these data as a mixture distribution of Poisson counts in which some patients are "cured" by the drug, whilst others experience varying levels of response but remain abnormal. A zero count for the post-drug PVC may indicate a "cure", or may represent a sampling zero from a patient with a mildly abnormal PVC count. The following model thus is assumed:

$x_i \sim$ Poisson(λ_i) for all patients
$y_i \sim$ Poisson($\beta\lambda_i$) for all *uncured* patients
P(cure) = θ

To eliminate nuisance parameters li, Farewell and Sprott use the conditional distribution of yi given ti = xi + yi. This is equivalent to a binomial likelihood for yi with denominator ti and probability p = b /(1+b) (see Cox and Hinkley, 1974 pp. 136-137 for further details of the conditional distribution for Poisson variables). Hence the final mixture model may be expressed as follows:

$$P(y_i = 0 \mid t_i) = \theta + (1 - \theta)(1 - p)^{t_i}$$
$$P(y_i \mid t_i) = (1 - \theta)\left(t_i! / (y_i!\,(t_i - y_i)!)\right)\left(p^{y_i}(1 - p)^{(t_i - y_i)}\right) \quad y_i = 1,2,\ldots,t_i$$

The BUGS code for this model is given below:

```
model
{
  for (i in 1 : N) {
    y[i] ~ dbin(P[state1[i]], t[i])
    state[i] ~ dbern(theta)
    state1[i] <- state[i] + 1
    t[i] <- x[i] + y[i]
    prop[i] <- P[state1[i]]
  }
  P[1] <- p
  P[2] <- 0
  logit(p) <- alpha
  alpha ~ dnorm(0,1.0E-4)
  beta <- exp(alpha)
  logit(theta) <- delta
  delta ~ dnorm(0, 1.0E-4)
}
```