BDA Assignment - Bayesian Inference

For one of the problem and models described below, you are required to implement python programs for **MCMC and variational inference** algorithms (without using PyMC3 or any packages which implements them but you are free to use packages such as scipy) to learn posterior distribution over the parameters (including hyperparameters). You need to provide samples from the posterior and various statistics like mean, median and variance. You are also required to compare your results of MCMC and VI routines with the MCMC and VI implementations using **PyMC3**. 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, mathematical expression for the lower bound used by VI for your problem and model, details of the sampler used, sampling process, proposal distribution, burn-in period etc.) , experimental setup, observations and results, and a comparison of the approaches for the chosen assignment problem. Data for the following problems can be found here
https://drive.google.com/drive/folders/1l7fGZG1fZ-m5i9pciOriWTAPXwOqO4R9?usp=sharing

Each question 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 problem should be taken on first come first serve basis (once 3 groups choose a problem, please choose another problem)
https://docs.google.com/spreadsheets/d/1LwTDYCuZ1q1wrm0Mg7sAojPLut8jQxKgmNEvPSn6wIo/edit?usp=sharing

Problem 1 : 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						

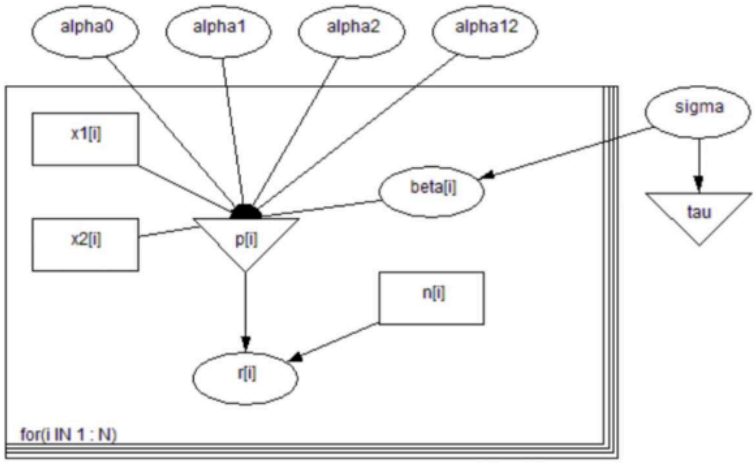
The model is essentially a random effects logistic, allowing for over-dispersion. If p_i is the probability of germination on the *i* th plate, we assume

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

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

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

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 2 : 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.

	PVC's per minute		
number (i)	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:

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

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)
}
```

Problem 3. Surgical: Institutional ranking

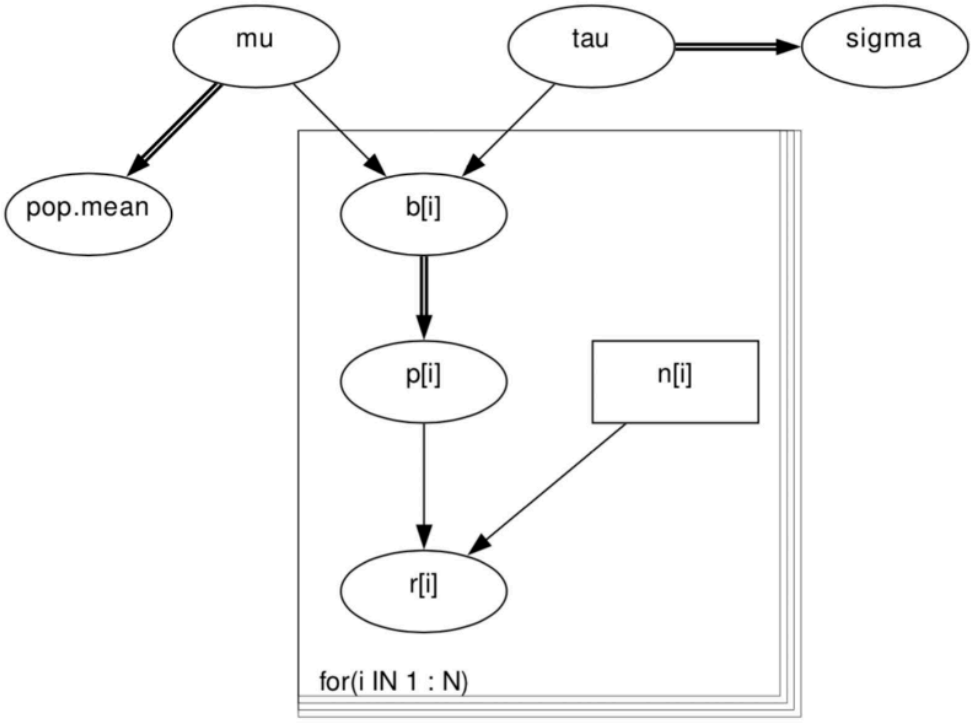
A realistic model for the surgical data is to assume that the failure rates across hospitals are *similar* in some way. This is equivalent to specifying a *random effects* model for the true failure probabilities p_i as follows:

$$\text{logit}(p_i) = b_i$$

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

Standard non-informative priors are then specified for the population mean probability of failure μ, and precision, τ.

Graphical model, pseudocode describing the model, and data for random effects surgical example:



model

{
for(i in 1 : N) {

b[i] ~ norm(mu,tau)

r[i] ~ bin(p[i],n[i])

logit(p[i]) <- b[i]
}

pop.mean <- exp(mu) / (1 + exp(mu)) mu ~ norm(0.0,1.0E-6)
sigma <- 1 / sqrt(tau)
tau ~ gamma(0.001,0.001)

Problem 4. Dyes: variance components model

Box and Tiao (1973) analyse data first presented by Davies (1967) concerning batch to batch variation in yields of dyestuff. The data (shown below) arise from a balanced experiment whereby the total product yield was determined for 5 samples from each of 6 randomly chosen batches of raw material.

Batch	Yield (in grams)				
1	1545	1440	1440	1520	1580
2	1540	1555	1490	1560	1495
3	1595	1550	1605	1510	1560
4	1445	1440	1595	1465	1545
5	1595	1630	1515	1635	1625
6	1520	1455	1450	1480	1445

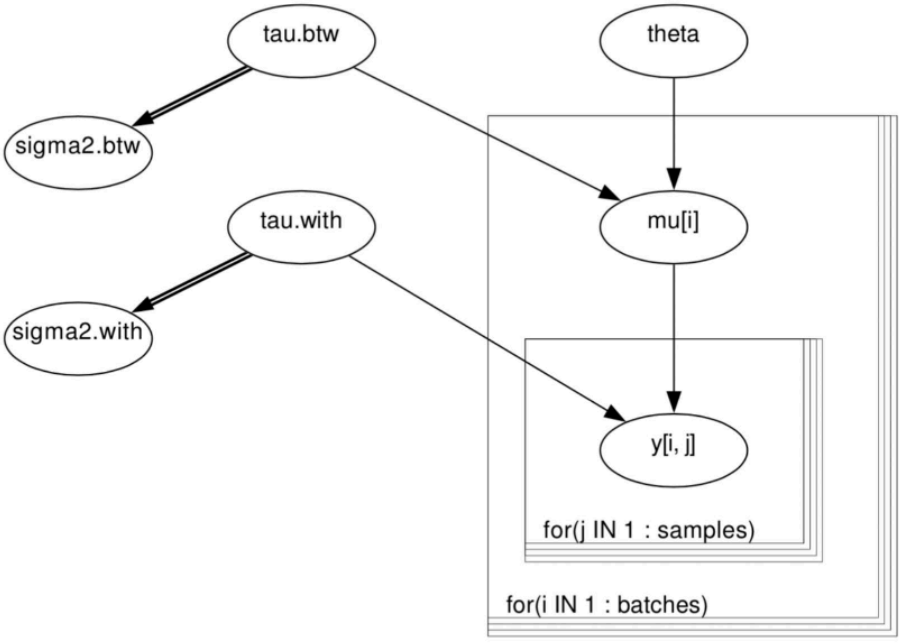
The object of the study was to determine the relative importance between batch variation versus variation due to sampling and analytic errors. On the assumption that the batches and samples vary independently, and contribute additively to the total error variance, we may assume the following model for dyestuff yield:

$y_{ij} \sim \text{Normal}(\mu_{.,}, \tau_{\text{within}})$

$\mu_i \sim Normal(\theta, \tau_{between})$

where y_{ij} is the yield for sample *j* of batch *i*, μ_i is the true yield for batch *i*, τ_{within} is the inverse of the within-batch variance σ^2_{within} (*i.e.* the variation due to sampling and analytic error), θ is the true average yield for all batches and $\tau_{between}$ is the inverse of the between-batch variance $s^2_{between}$. The total variation in product yield is thus $\sigma^2_{total} = \sigma^2_{within} + \sigma^2_{between}$ and the relative contributions of each component to the total variance are $f_{within} = \sigma^2_{within} / \sigma^2_{total}$ and $f_{between} = \sigma^2_{between} / \sigma^2_{total}$. We assume standard non-informative priors for θ, τ_{within} and $\tau_{between}$.

Graphical model, pseudocode describing the model, and data *for dyes example is below*



{

```
for(i in 1 : batches) {

m[i] ~ norm(theta, tau.btw) for(j in 1 : samples) {

y[i , j] ~ norm(m[i], tau.with) }

}
sigma2.with <- 1 / tau.with

sigma2.btw <- 1 / tau.btw

tau.with ~ gamma(0.001, 0.001)

tau.btw ~ gamma(0.001, 0.001)

theta ~ norm(0.0, 1.0E-10)

}
```

Problem 5. LSAT: item response

Law School Aptitude Test (LSAT) is a 5-item multiple choice test; students score 1 on each item for the correct answer and 0 otherwise, giving R = 32 possible response patterns.Boch and Lieberman (1970) present data on LSAT for N = 1000 students, part

Pattern index	Item response pattern	Freq (m)
1	0 0 0 0 0	3
2	0 0 0 0 1	6
3	0 0 0 1 0	2
.	. . . . .	.
.	. . . . .	.
.	. . . . .	.
30	1 1 1 0 1	61
31	1 1 1 1 0	28
32	1 1 1 1 1	298

of which is shown below.

The above data may be analysed using the one-parameter Rasch model. The probability p_{jk} that student *j* responds correctly to item *k* is assumed to follow a logistic function parameterized by an `item difficulty' or threshold parameter α_k and a latent variable θ_j representing the student's underlying ability. The ability parameters are assumed to have a Normal distribution in the population of students.

The model and data is described below, prior is defined over thetai, alpha and beta.

$$
\begin{aligned}
r_{i,j} &\sim \text{Bernoulli}(p_{i,j}) \qquad i = 1, \ldots, 1000; j = 1, \ldots, 5 \\
\text{logit}(p_{i,j}) &= \beta\theta_i - \alpha_j \\
\theta_i &\sim \text{Normal}(0, 1) \\
\alpha_j &\sim \text{Normal}(0, 100) \\
\beta &\sim \text{Flat}(0, \infty),
\end{aligned}
$$

```
# Rasch model
   for (j in 1 : N) {
      for (k in 1 : T) {
         logit(p[j, k]) <- beta * theta[j] - alpha[k]
         r[j, k] ~ dbern(p[j, k])
      }
      theta[j] ~ dnorm(0, 1)
   }
# Priors
   for (k in 1 : T) {
      alpha[k] ~ dnorm(0, 0.0001)
      a[k] <- alpha[k] - mean(alpha[])
   }
   beta ~ dnorm(0,0.0001) I(0, )
}
```

Problem 6. Equiv: bioequivalence in a cross-over trial

The table below shows some data from a two-treatment, two-period crossover trial to compare 2

Subject *i*	Sequence	*seq*	Period 1	T_{i1}	Period 2	T_{i2}
1	AB	1	1.40	1	1.65	2
2	AB	1	1.64	1	1.57	2
3	BA	-1	1.44	2	1.58	1
....						
8	AB	1	1.25	1	1.44	2
9	BA	-1	1.25	2	1.39	1
10	BA	-1	1.30	2	1.52	1

tablets A and B, with 10 subjects.

The response Y_{ik} from the *i* th subject (*i* = 1,...,10) in the *k* th period (*k* = 1,2) is assumed to be of the form

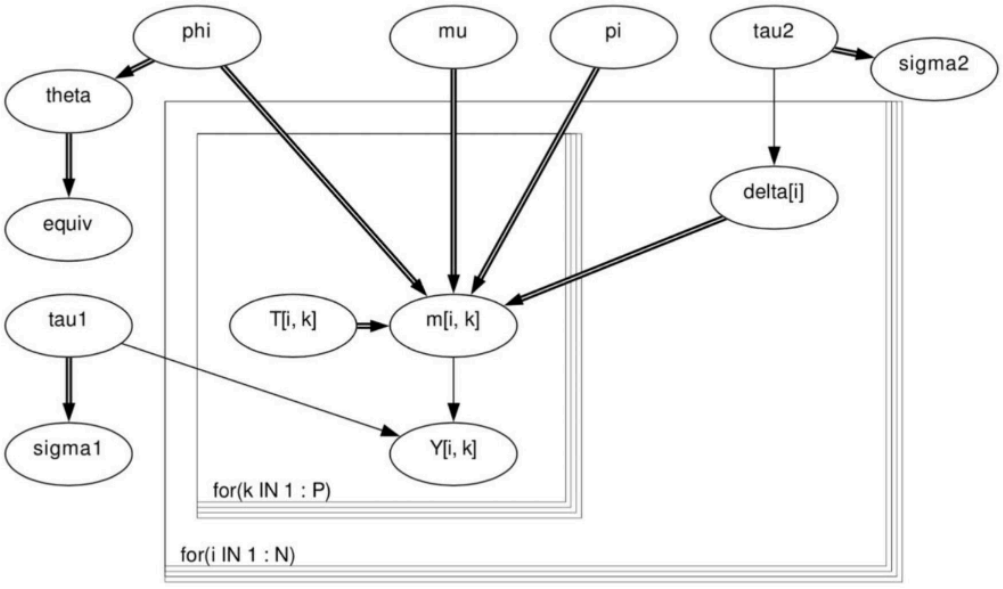
Treatment responses are modelled as

$$
\begin{aligned}
y_{i,j} &\sim \text{Normal}(m_{i,j}, \sigma_1) \qquad i = 1, \ldots, 10; j = 1, 2 \\
m_{i,j} &= \mu + (-1)^{T_{i,j}-1}\phi/2 + (-1)^{j-1}\pi/2 + \delta_i \\
\delta_i &\sim \text{Normal}(0, \sigma_2) \\
\mu, \phi, \pi &\sim \text{Normal}(0, 1000) \\
\sigma_1^2, \sigma_2^2 &\sim \text{InverseGamma}(0.001, 0.001)
\end{aligned}
$$

where $y_{i,j}$ is the response for patient i in period j; and $T_{i,j} = 1, 2$ is the treatment received.

μ, ϕ, π are the overall mean, treatment and period effects respectively, and δ_i represents the random effect for subject *i*.

The graph of this model, pseudocode describing the model and data are given below




```
{
for( k in 1 : P ) {

for( i in 1 : N ) {
Y[i , k] ~ norm(m[i , k], tau1)
m[i , k] <- mu + sign[T[i , k]] * phi / 2 + sign[k] * pi / 2 + delta[i]

T[i , k] <- group[i] * (k - 1.5) + 1.5

}

}

for( i in 1 : N ) {
```

```
delta[i] ~ norm(0.0, tau2) }

tau1 ~ gamma(0.001, 0.001) sigma1 <- 1 / sqrt(tau1)

tau2 ~ gamma(0.001, 0.001) sigma2 <- 1 / sqrt(tau2)

mu ~ norm(0.0, 1.0E-6)
phi ~ norm(0.0, 1.0E-6)

pi ~ norm(0.0, 1.0E-6)
theta <- exp(phi)


}
```