Before starting, we should calculate the posterior.
For any $X \sim N(\mu, \sigma^2 I)$, we can view $X_1, X_2$ are independent draws from $N(\mu_1; \sigma^2), N(\mu_2, \sigma^2)$ respectively. We have that

$$p(X) \propto \sigma^{-2} \exp\left(-\frac{(X_1-\mu_1)^2+(X_2-\mu_2)^2}{2\sigma^2}\right) \propto$$

$$\propto \sigma^{-2} \exp\left(-\frac{X_1^2 - 2X_1\mu_1 + \mu_1^2 + X_2^2 - 2X_2\mu_2 + \mu_2^2}{2\sigma^2}\right)$$

Now, let's $Y_{ijk}$ be the $k$-th element of sample $j$ from group $i$. Thus, since our prior has $\pi(\theta) \propto \sigma^{-2}$, it follows for $n = \sum_{i=1}^{4} n_i$ that

$$p(\sigma^2, \lambda, \tau, \mu_1, \mu_2, \gamma_1, \gamma_2 \mid Y) \propto$$

$$\propto \sigma^{-2(n+1)} \exp\left(-\sum_{j=1}^{n_1} \frac{Y_{1j1}^2 - 2Y_{1j1}\mu_1 + \mu_1^2 + Y_{1j2}^2 - 2Y_{1j2}\mu_2 + \mu_2^2}{2\sigma^2}\right) \cdot$$

$$\exp\left(-\sum_{j=1}^{n_2} \frac{Y_{2j1}^2 - 2Y_{2j1}\gamma_1 + \gamma_1^2 + Y_{2j2}^2 - 2Y_{2j2}\gamma_2 + \gamma_2^2}{2\sigma^2}\right) \cdot$$

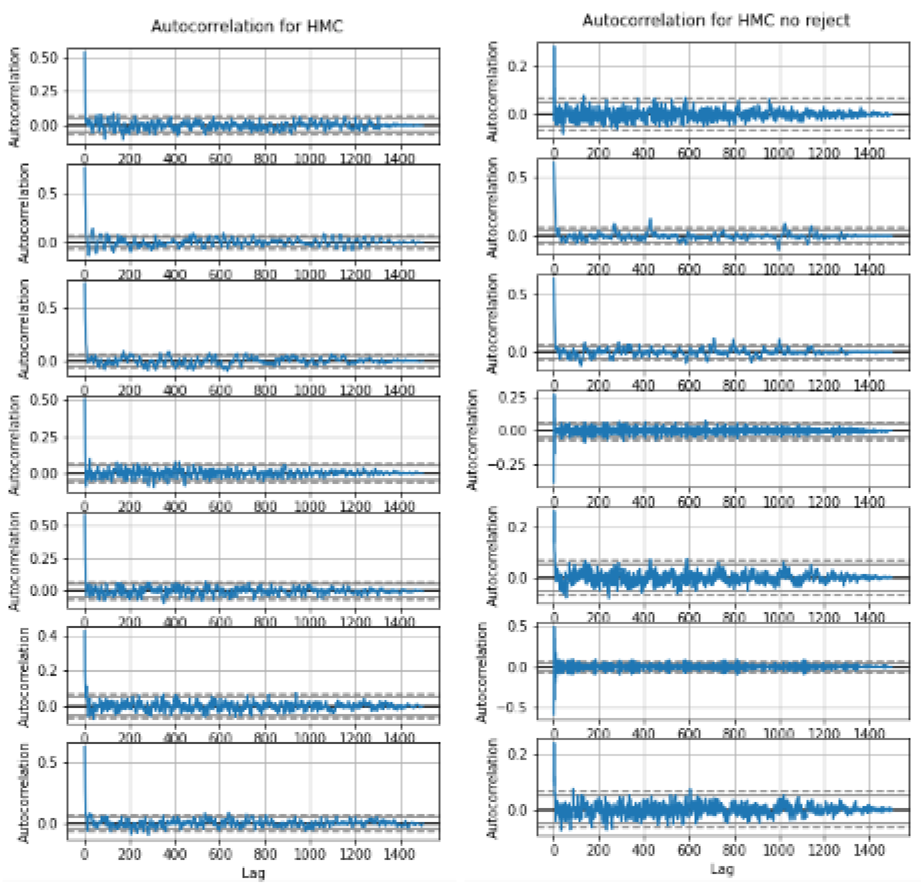
$$\cdot \exp\left(-\sum_{j=1}^{n_3} \sum_{k=1}^{2} \frac{Y_{3jk}^2 - 2Y_{3jk}(\lambda\mu_k + (1-\lambda)\gamma_k) + (\lambda\mu_k + (1-\lambda)\gamma_k)^2}{2\sigma^2}\right) \cdot$$

$$\cdot \exp\left(-\sum_{j=1}^{n_4} \sum_{k=1}^{2} \frac{Y_{3jk}^2 - 2Y_{3jk}(\tau\mu_k + (1-\tau)\gamma_k) + (\tau\mu_k + (1-\tau)\gamma_k)^2}{2\sigma^2}\right)$$

1| Hamiltonian Monte Carlo (HMC)

I used two models for HMC. (1) HMC without rejection correction, and (2) HMC with rejection correction by Metropolis Hasting rule. The reason I used these two is because I wanted to compare easier and more complex versions of HMC. Below I will

**<u>Please see the Codes Attached</u>**



2) Metropolis-Hasting

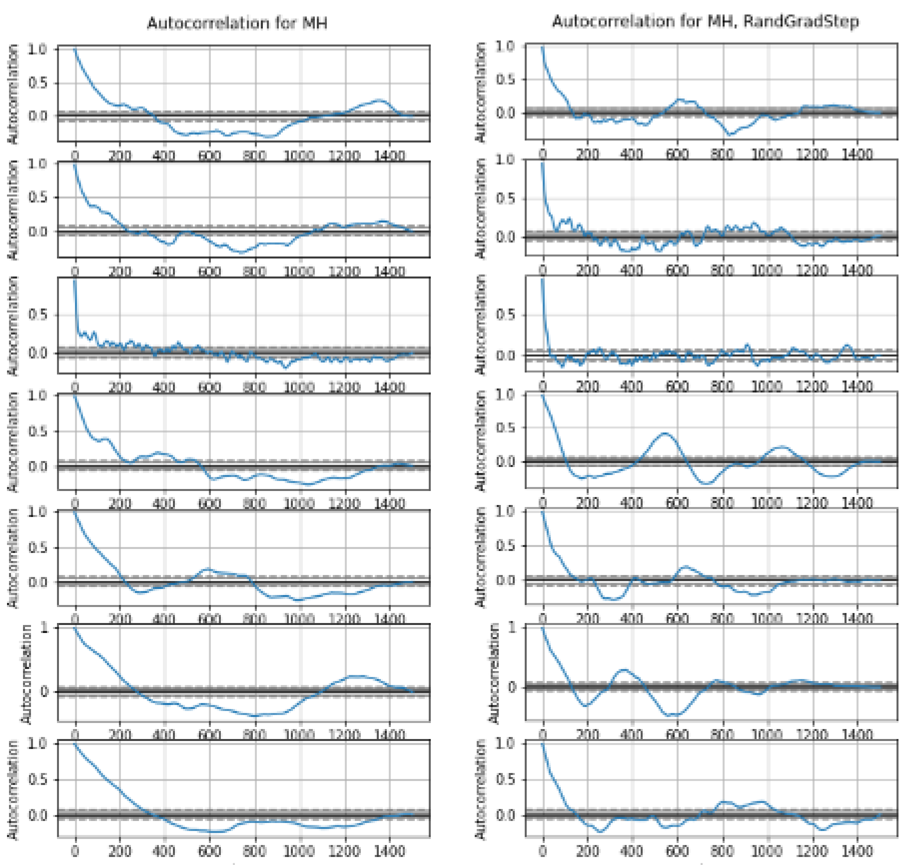
For the Metropolis-Hasting I also used two methods. (1) For the first method, the proposal is to move $\theta$ to $\theta + \epsilon Z$ where $Z \sim N(0, I)$ where $I$ is the $J \times J$ identity matrix, and $\epsilon$ is a step size. (2) For the second method, the proposal is to move from $\theta$ to $\theta + \delta D g(\theta) + \epsilon Z$ where $D$ is +1 or -1 with probability 0.5 each, $g(\theta)$ is the gradient of the log-posterior density, $Z \sim N(0, I)$ and $\delta, \epsilon$ are step sizes.

As we know from the MH, for a current point $\theta$, we propose a new $\theta'$ under some rule. We the calculate the acceptance ratio.

$$A(\theta, \theta') = \frac{P(\theta')\, p(\theta' | \theta)}{P(\theta)\, p(\theta | \theta')}$$

where $P(\theta)$ is the posterior distribution for $\theta$ and $p(\theta' | \theta)$ is the density for proposing $\theta'$ given we are currently at $\theta$. Since we are taking a ration $P(\theta')/P(\theta)$ we can ignore normalizing constants. For the (1) case since the Gaussian $Z$ is symmetric, we have ~~that~~ $p(\theta'|\theta) = p(\theta|\theta')$. For the (2) we have that $p(\theta'|\theta) = \frac{1}{2} p_{N(0,I)}(Z) + \frac{1}{2} p_{N(0,I)}(Z_{alt})$, where we define $Z_{alt} = \frac{\epsilon Z + 2D\delta g(\theta)}{\epsilon}$ which is the step we need to take to arrive at $\theta'$ when $D$ was the other direction. A similar calculation can be made for $p(\theta|\theta')$. Below you could

**Please see the Codes Attached**



## 3) Gibbs Sampler

We take Gibbs sampling where one of $[\sigma^2]$, $[\lambda, \tau]$ or $[\mu_1, \mu_2, \gamma_1, \gamma_2]$ is updated in each step.

Let us try to find the appropriate posterior distributions.

$$P(\sigma^2 \mid \lambda, \tau, \mu_1, \mu_2, \gamma_1, \gamma_2, Y) \propto \sigma^{-2(n+1)} \exp\left(-\frac{1}{2\sigma^2} \sum_{ijk} \left(Y_{ijk} - \mathbb{E}[Y_{ijk}]\right)^2\right)$$

Thus, we have that $\sigma^2$ is Inverse Gamma with parameters $\alpha = n$, $\beta = \frac{1}{2} \sum_{ijk} (Y_{ijk} - \mathbb{E}[Y_{ijk}])^2$. $\lambda, \tau$ are independent, so without loss of generality the conditional distribution of $\lambda$ is

$$P(\lambda \mid \sigma^2, \mu_1, \mu_2, \gamma_1, \gamma_2, Y) \propto$$

$$\propto \exp\Bigg( - \sum_{j=1}^{n_3} \frac{-2Y_{3j1}(\lambda w_{1} + (1-\lambda)\gamma_1) + (\lambda w_{1} + (1-\lambda)\gamma_1)^2}{2\sigma^2}$$

$$- \sum_{j=1}^{n_3} \frac{-2Y_{3j2}(\lambda w_{2} + (1-\lambda)\gamma_2) + (\lambda w_{2} + (1-\lambda)\gamma_2)^2}{2\sigma^2}\Bigg) \mathbb{I}(\lambda \in [0,1])$$

$$\propto \exp\Bigg( \frac{\lambda}{2\sigma^2} \sum_{j=1}^{n_3} \sum_{k=1}^{2} [2Y_{3jk}(w_{k} - \gamma_k) + 2\gamma_k^2 - 2w_{k}\gamma_k]$$

$$- \frac{\lambda^2}{2\sigma^2} \sum_{j=1}^{n_3} \sum_{k=1}^{2} (w_{k} - \gamma_k)^2 \Bigg) \mathbb{I}(\lambda \in [0,1])$$

which is a truncated normal distribution.

Finally, note that $w_1, \gamma_1 \perp w_2, \gamma_2$. Thus,

$$P(w_1 \mid \lambda, \tau, \sigma^2, Y) \propto \exp\Bigg( - \sum_{j=1}^{n_1} \frac{-2Y_{1j1} w_1 + w_1^2}{2\sigma^2}$$

$$- \sum_{j=1}^{n_3} \frac{-2Y_{3j1}(\lambda w_1 + (1-\lambda)\gamma_1) + (\lambda w_1 + (1-\lambda)\gamma_1)^2}{2\sigma^2}\Bigg)$$

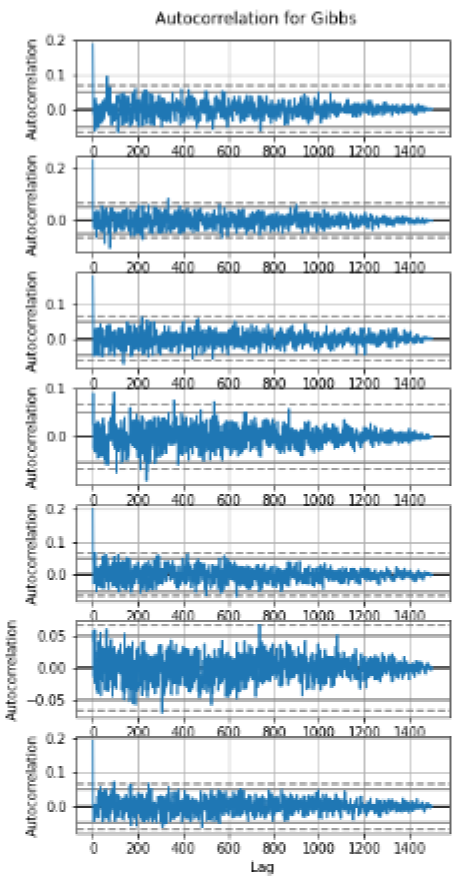
$$- \sum_{j=1}^{n_4} \frac{-2Y_{4j1}(\tau w_1 + (1-\tau)\gamma_1) + (\tau w_1 + (1-\tau)\gamma_1)^2}{2\sigma^2}$$

$$\propto \exp\Bigg( \frac{w_1}{2\sigma^2} \Bigg[ \sum_{j=1}^{n} 2Y_{1j1} + \sum_{j=1}^{n_3} 2Y_{3j1}\lambda - 2\lambda(1-\lambda)\gamma_1 +$$

$$+ \sum_{j=1}^{n_4} 2Y_{4j1}\tau - 2\tau(1-\tau)\gamma_1 \Bigg] - \frac{w_1^2}{2\sigma^2}[n_1 + \lambda^2 n_3 + \tau^2 n_4] \Bigg)$$

which is once again a normal distribution. these distributions are all easy to sample from, so we can use Gibbs sampling to generate samples from the posterior.

**<u>Please see the Codes Attached</u>**



4) Importance sampling

For the Importance sampling we have
S samples $\theta_1 \ldots \theta_S$ from the proposed density $g(\theta)$, sample of a given size from true density $w(\theta) = \frac{p(\theta|y)}{g(\theta)}$ — weight, and we

$$E[f(x)] = \int f(\theta)\, p(\theta)\, d\theta = \int f(\theta) \frac{p(\theta)}{q(\theta)}\, q(\theta)\, d\theta \approx$$

$$\approx \frac{1}{n} \sum_i f(\theta_i) \frac{p(\theta_i)}{q(\theta_i)}.$$

My codes for Importance sampling are below; However, I am getting an ~~error~~ error, so I need a bit more time to fix it. But I completed it (with errors).

**Please see the Codes Attached**

Overall, HMC was really good. It was really quick and consistent. HMC without rejection ended up having the largest error. Also note, that HMC was with leapfrog algorithm which requires many full scans of the data. If the data is large, it can be expensive.

MH and Gibbs were also quite good even thoug they are simpler. However MH may suffer if the initial value is very far from the probability region.

I would say HMC and Gibbs were the best for their consistency, error, and speed.

In terms of autocorrelations, we want them to be as close to 0 as possible. Overall, all the method are not too far from 0, in term of autocorrelation, however, HMC method (2) has autocorrelation that is closer to 0 and varies less.