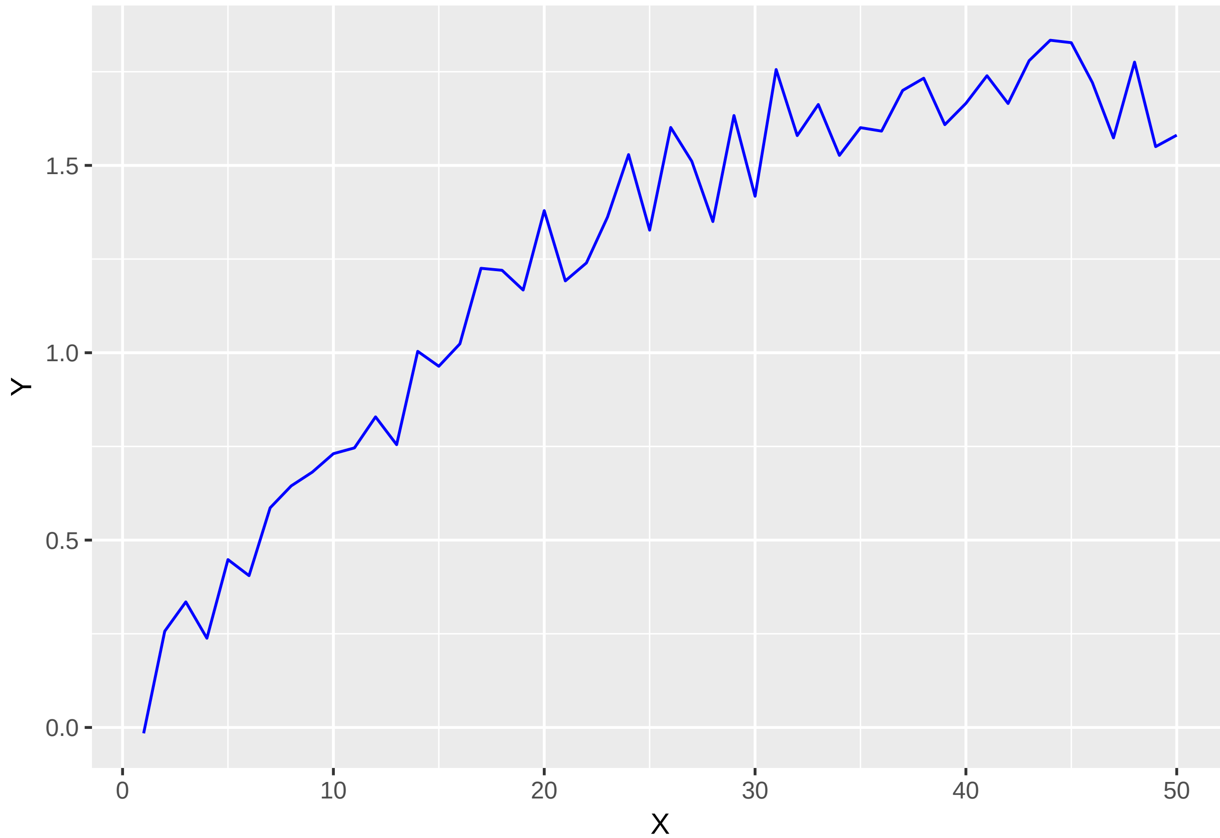
Gibbs Sampling

##. Read Data and Plot X vs Y



I think a polynomial regression would be a good fit to the data. A curve would perfecly catch the dependence between X and Y.

##. Calculate Likelyhood and prior

Given :

$$Y_i \sim \mathcal{N}(\mu_i, variance = 0.2)$$
$$p(\mu_1) = 1$$
$$p(\mu_{i+1}|\mu_i) \sim \mathcal{N}(\mu_i, 0.2), i = 1, 2, 3..., n-1$$

Where prior is $p(\mu_1)$ and $p(\mu_{i+1}|\mu_i)$

$$p(\mu) = p(\mu_1).p(\mu_2|\mu_1).p(\mu_3|\mu_2)...p(\mu_n|\mu_{n-1})$$
$$p(\mu) = k.e^{-\frac{1}{2.\sigma^2}.\sum_{i=1}^{n-1}(\mu_{i+1}-\mu i)^2}$$

Calculating likelyhood :

$$p(Y|\vec{\mu}) = p(Y_1|\vec{\mu_1}).p(Y_2|\vec{\mu_2}).p(Y_3|\vec{\mu_3})...p(Y_n|\vec{\mu_n})$$
$$p(Y|\vec{\mu}) = \prod_{i=1}^{n} p(Y_i|\vec{\mu_i})$$
$$p(Y|\vec{\mu}) = k.e^{-\frac{1}{2.\sigma^2}.\sum_{i=1}^{n}(Y_i-\mu i)^2}$$

Calculating posterior and conditional distributions

$$P(\mu) \propto P(Y|\mu).P(\mu)$$
$$P(\mu) \propto k.e^{-\frac{1}{2.\sigma^2}.\sum_{i=1}^{n}(Y_i-\mu i)^2}.e^{-\frac{1}{2.\sigma^2}.\sum_{i=1}^{n-1}(\mu_{i+1}-\mu i)^2}$$
$$P(\mu) \propto e^{-\frac{1}{2.\sigma^2}(\sum_{i=1}^{n}(Y_i-\mu i)^2+\sum_{i=1}^{n-1}(\mu_{i+1}-\mu i)^2)}$$

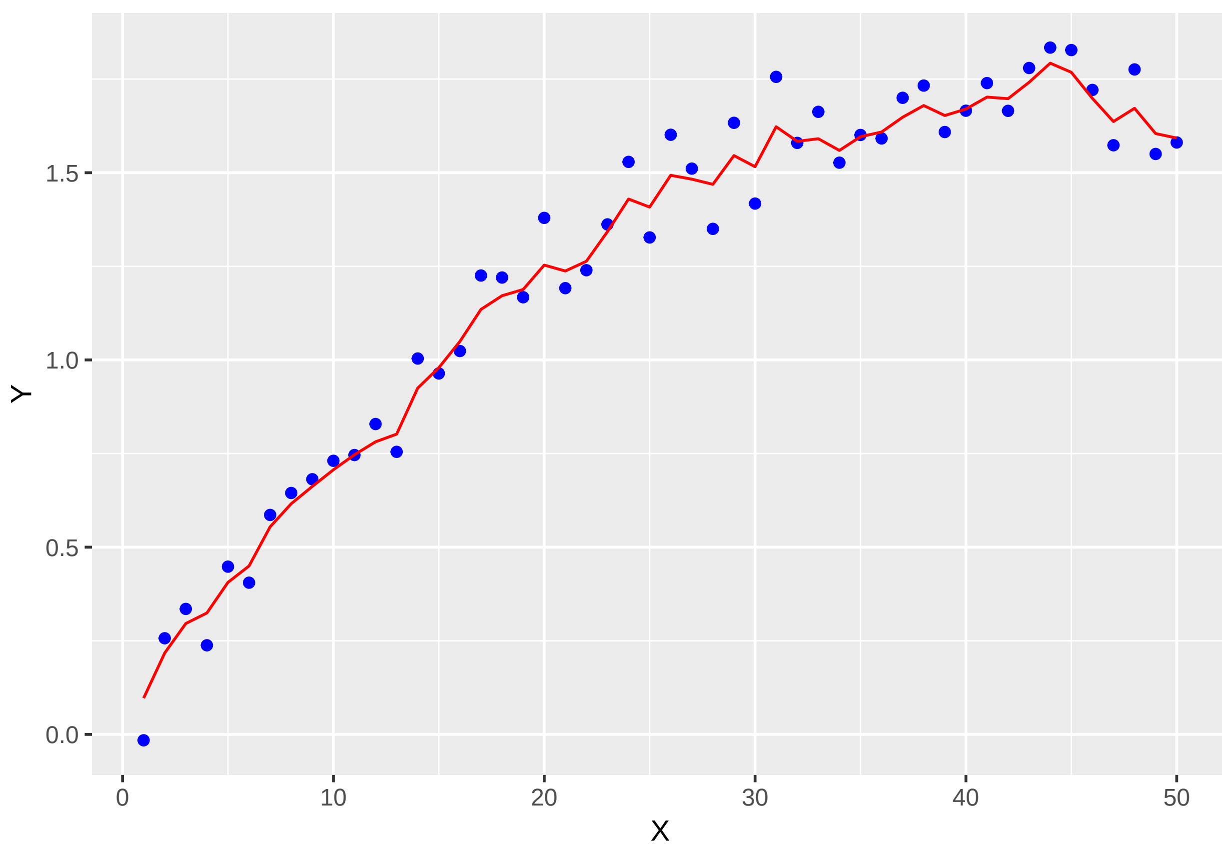
Using Hint A, B and C we can get conditional distribution as:

$$P(\mu_1|\mu_{-1},Y) = e^{-\frac{1}{2.\sigma^2}(\mu_1-(\frac{\mu_2+y_1}{2}))^2}$$
$$P(\mu_i|\mu_{-i},Y)) = e^{-\frac{1}{2.\sigma^2}(\mu_i-(\frac{\mu_{i-1}+\mu_{i+1}+y_n}{3}))^2}$$
$$P(\mu_n|\mu_{-n},Y)) = e^{-\frac{1}{2.\sigma^2}(\mu_n-(\frac{\mu_{n-1}+y_n}{2}))^2}$$

Each of the above distribution is normally distributed with below properties :

$$\mu_1 \sim \mathcal{N}(\frac{\mu_2+Y_1}{2}, \sigma^2 = 0.2)$$
$$\mu_i \sim \mathcal{N}(\frac{\mu_{i-1}+\mu i+1+Y_i}{3}, \sigma^2 = 0.2)$$
$$\mu_n \sim \mathcal{N}(\frac{\mu_{n-1}+Y_n}{2}, \sigma^2 = 0.2)$$

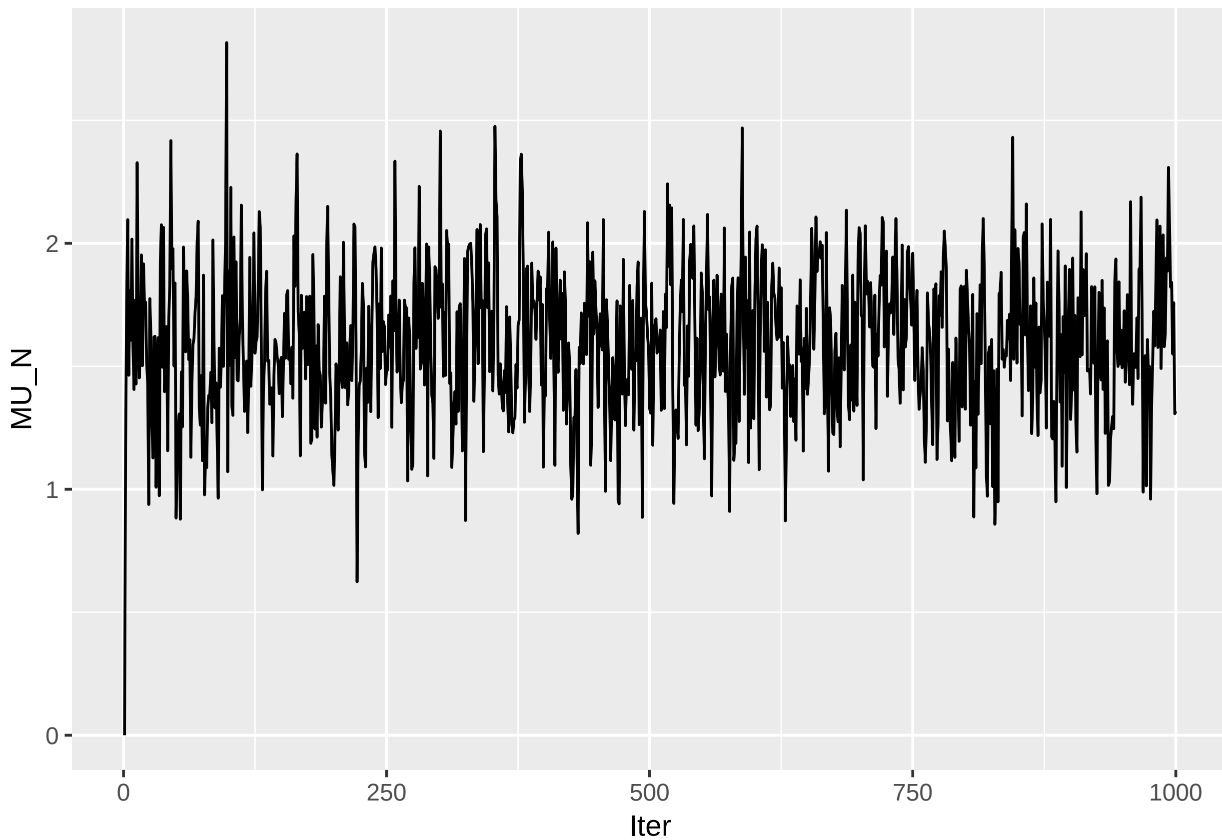
##. Gibbs Sampler



Yes it looks like the line fits the data quit well removing much of the noise.

I think the expected value of mu catches the underlying dependence between X and Y.

##. Trace Plot



The mu value gets close to the actual distribution of Y after the first iteration. For μ_n the update in the first iteration ends up being the average of Y_n and mu_n-1(which is close to the average of the previous two points of Y), so the mu_n for the first iteration ends up being close to the average of the points Y_i and Y_{i-1}. So the burn in period would be just one iteration.

The value of mu keeps fluctuating. Since we are generating a random number using rnorm with a small standard deviation in each iteration, the value of mu stays close to the average of the points Y_n and Y_{n-1}.