Problem 1

A. Derivations:-

Mathematical derivation for bivariate Gaussian distribution $N_2(\mu, \Sigma)$:-

- We can find generalized multivariate Gaussian distribution and reduce bivariate Gaussian distribution.

Let $X \sim N(\mu, \Sigma)$

where $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

s.t $x_1 \in R^p$, $x_2 \in R^q$ $\Rightarrow x \in R^N$ where $N = p+q$

$\Sigma_{11} \in R^{p \times q}$ $\Sigma_{22} \in R^{q \times q}$ $\Sigma_{12} \in R^{p \times q}$ and $\Sigma_{21}^T = \Sigma_{12}$ [Symmetric property of Covariance matrix]

So $p(X) = p(x_1, x_2) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$ —(3)

Let $Q(x_1, x_2) = (x-\mu)^T \Sigma^{-1} (x-\mu)$ — (4)

$= \begin{bmatrix} (x_1-\mu_1)^T & (x_2-\mu_2)^T \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$

Let $\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$ —(6)

Using block matrix multiplication

$Q(x_1, x_2) = \begin{bmatrix} (x_1-\mu_1)^T \Sigma^{11} + (x_2-\mu_2)^T \Sigma^{21}, & (x_1-\mu_1)^T \Sigma^{12} + (x_2-\mu_2)^T \Sigma^{22} \end{bmatrix} * \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{bmatrix}$

$$= (x_1-\mu_1)^T \Sigma^{11} (x_1-\mu_1) + (x_2-\mu_2)^T \Sigma^{21} (x_1-\mu_1) + (x_1-\mu_1)^T \Sigma^{12} (x_2-\mu_2) + (x_2-\mu_2)^T \Sigma^{22} (x_2-\mu_2)$$

using $(\Sigma^{11})^T = \Sigma^{12}$

$$Q(x_1, x_2) = (x_1-\mu_1)^T \Sigma^{11} (x_1-\mu_1) + 2(x_1-\mu_1)^T \Sigma^{12} (x_2-\mu_2) + (x_2-\mu_2)^T \Sigma^{22} (x_2-\mu_2) \quad -(8)$$

Since we assumed that

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$$

So $\Sigma^{11} = \Sigma_{11}^{-1} + \Sigma_{11}^{-1} \Sigma_{12} \left(\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12} \right)^{-1} \Sigma_{12}^T \Sigma_{11}^{-1}$

$\Sigma^{22} = \Sigma_{22}^{-1} + \Sigma_{22}^{-1} \Sigma_{12}^T \left(\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T \right)^{-1} \Sigma_{12} \Sigma_{22}^{-1}$

$\Sigma^{12} = -\Sigma_{11}^{-1} \Sigma_{12} \left(\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12} \right)^{-1} = (\Sigma^{21})^T$

So using above values in eq-(8)

$$Q(x_1, x_2) = (x_1-\mu_1)^T (\Sigma_{11}^{-1}) (x_1-\mu_1) + \left[x_2 - \mu_2 - \Sigma_{12}^T \Sigma_{11}^{-1} (x_1-\mu_1) \right]^T * \left(\Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12} \right)^{-1} \left[x_2 - \mu_2 - \Sigma_{12}^T \Sigma_{11}^{-1} (x_1-\mu_1) \right]$$

or $Q(x_1, x_2) = Q_1(x_1) + Q_2(x_1, x_2)$

Heuristically, the addition in $Q(x_1, x_2)$ become product when plugged back to original formula of multivariate Gaussian distribution.

Thus, divide the expression by $p(x_i)$, we will end up term produced by $Q_2(x_1, x_2)$. So by using heuristic,

$$X_2/X_1 \sim N\left(u_2 + \Sigma_{12}^T \Sigma_{11}^{-1} (x_1 - u_1),\ \Sigma_{22} - \Sigma_{12}^T \Sigma_{11}^{-1} \Sigma_{12}\right)$$

Now for the given problem, if $x_1 \in R^1$, $x_2 \in R$, $\Sigma_{12}, \Sigma_{11}, \Sigma_{22}, \Sigma_{21} \in R^1$ then the conditional probability can be written as

$$X_2/X_1 \sim N\left(u_2 + \frac{\Sigma_{12}}{\Sigma_{11}} (x_1 - u_1),\ \frac{\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2}{\Sigma_{11}}\right)$$

and

$$X_1/X_2 \sim N\left(u_1 + \frac{\Sigma_{12}}{\Sigma_{22}} (x_2 - u_2),\ \frac{\Sigma_{11}\Sigma_{22} - \Sigma_{12}^2}{\Sigma_{22}}\right)$$

Problem(B). Implemented in Solution_Prob_1.py

Problem(C) : For a=0 [Please run Solution_Prob_1.py with a=0 for results]

X1 vs X2 plot is more spreaded compared to a=0.99. As a tends to 1, x1 vs x2 probability density spread will reduce in graph and converges to a single point (in plot).

Fig.1 represents x1 vs x2 along the line between the last 2 latest points.

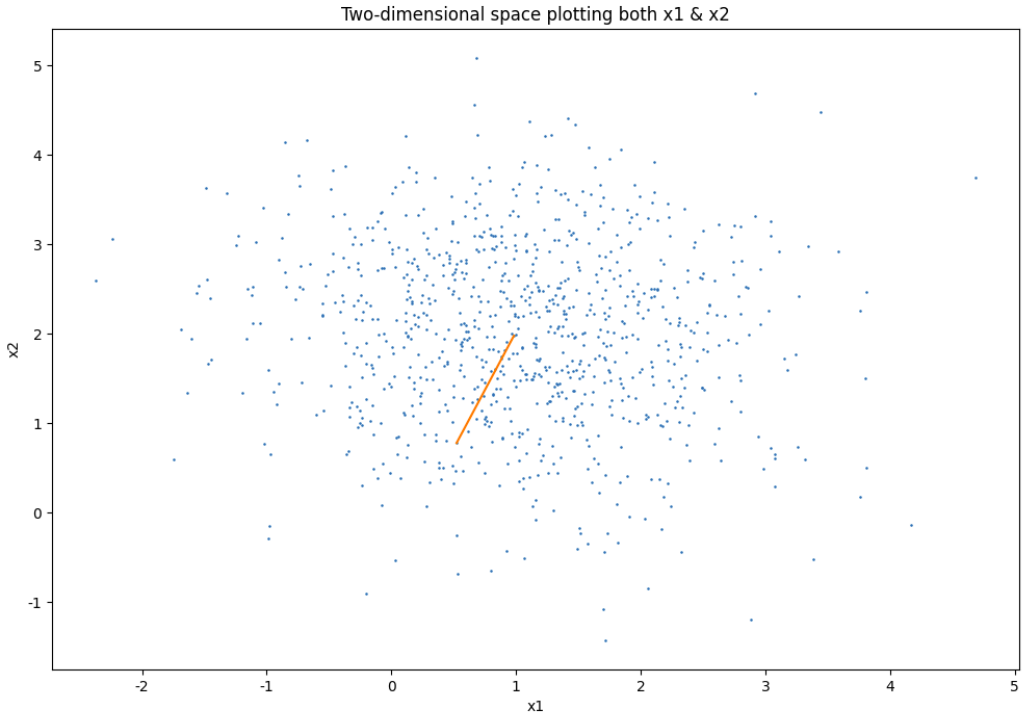


Fig 1

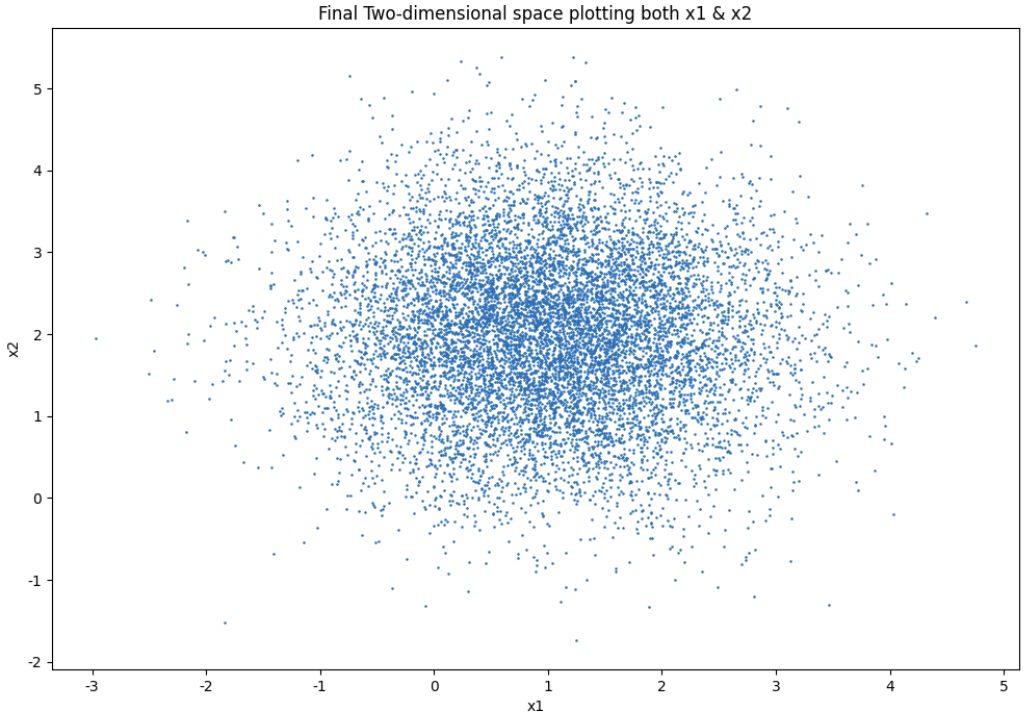


Fig. 2 [Final plot for 10,000 points]

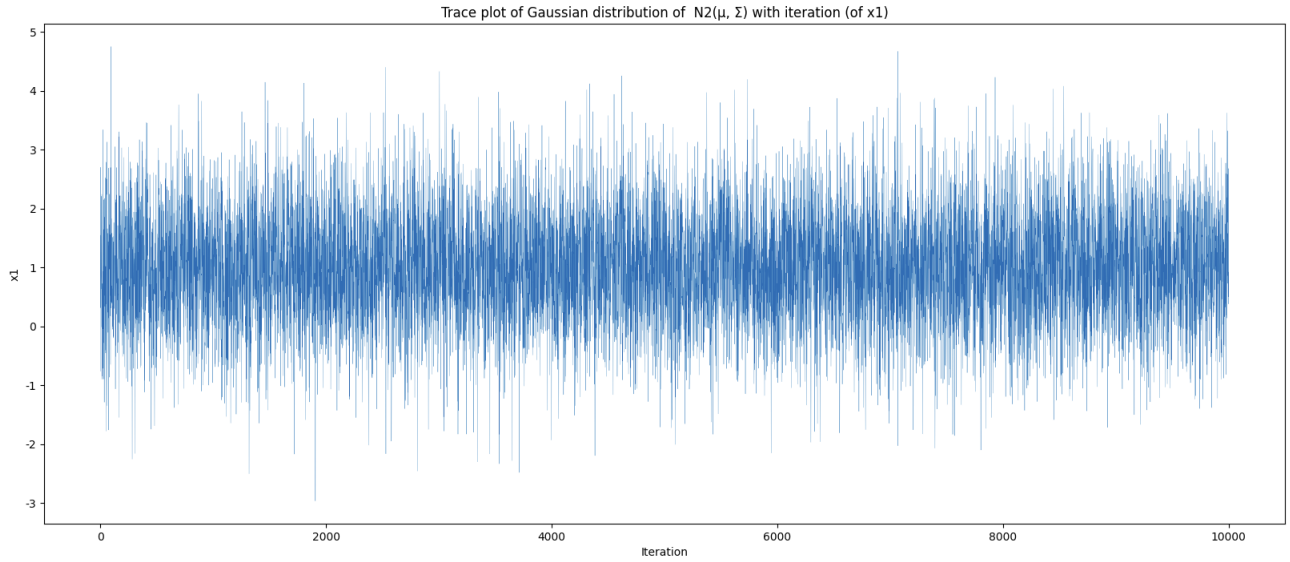


Fig 3. [Trace plot of x1 w.r.t iterations, for a=0]

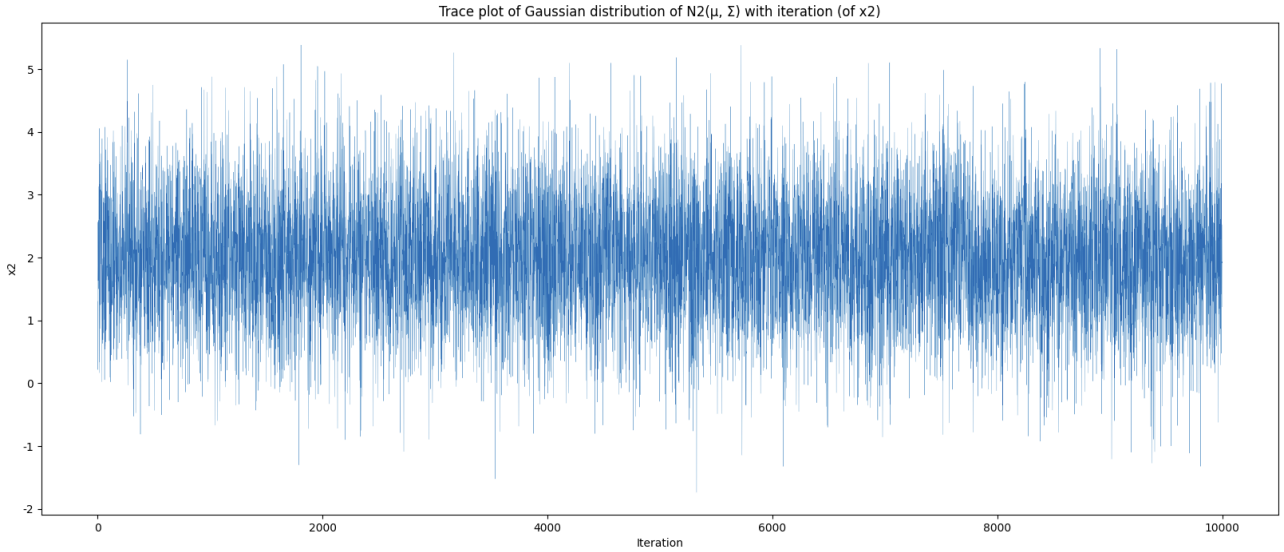


Fig 4. [Trace plot of x2 w.r.t iterations for a=0]

In this algorithm, the samples are produced from marginal distributions for both x1 and x2. Places in the trace plot represent regions of the density in which the algorithm is simulated frequency. Except the blue region (all graph), other regions (area) will rarely be visited by the algorithm.

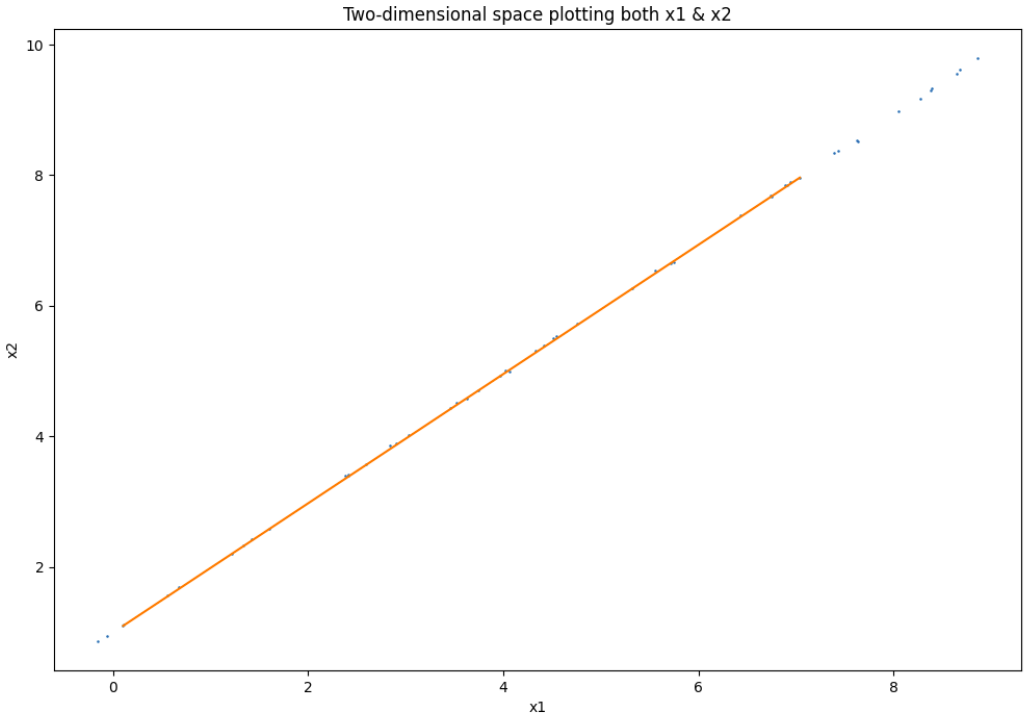
For a=0, traceplot values (x1 and x2) quickly converge (or already converged) compare to a=0.99 which takes more iterations to converge.

From the above traceplot, since the both marginal probability x1 and x2 are accuracy, we don't need to burn in or discard any initial iterations.

Problem D:

Same program run for 10,000 iterations with a=0.99, here are the graphs of the following:-

X1 vs x2 iterations graph:



In the above graph, the red color line is drawn between the last 2 points in x1 vs x2 value. All the points are allocated in the small region as the **a's** value is increasing. With lesser value of **a,** the points in x1 vs x2 plot are more spreaded. If a=1, the x1 vs x2 marginal probability always forms a single point and trace plot of x1 and x2 will form a straight line.

After 10,000 iterations, (final graph of x1 vs x2):

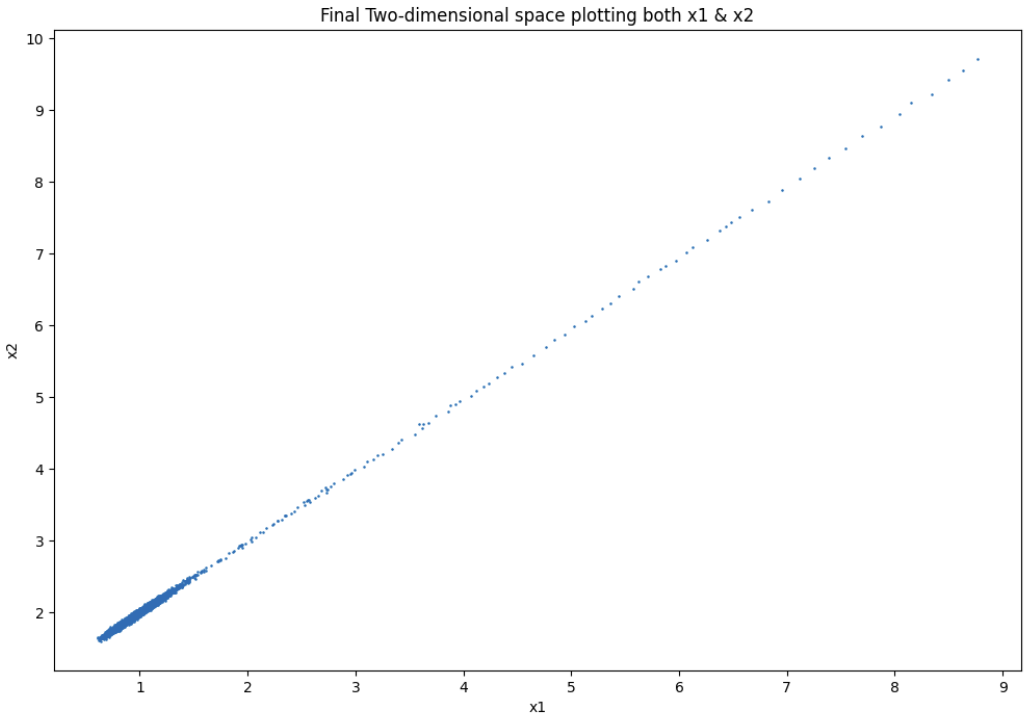


Fig. 5 [Final plot for 10,000 points plot, for a=0.99]

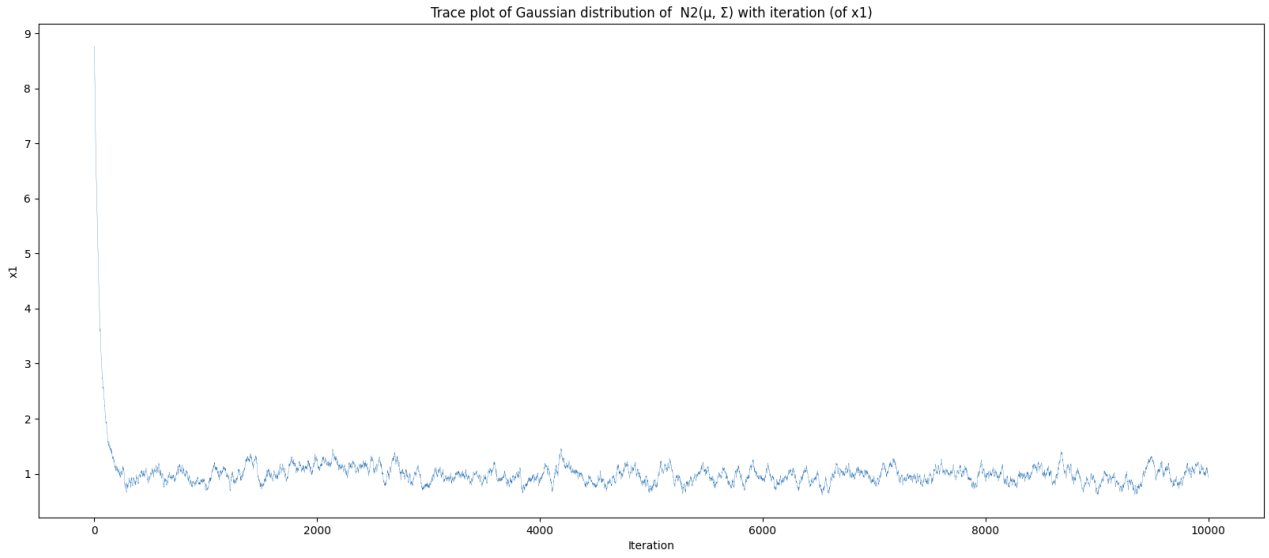


Fig 6. [Trace plot of x1 w.r.t iterations, for a=0.99]

Trace plot of x2 vs iterations:

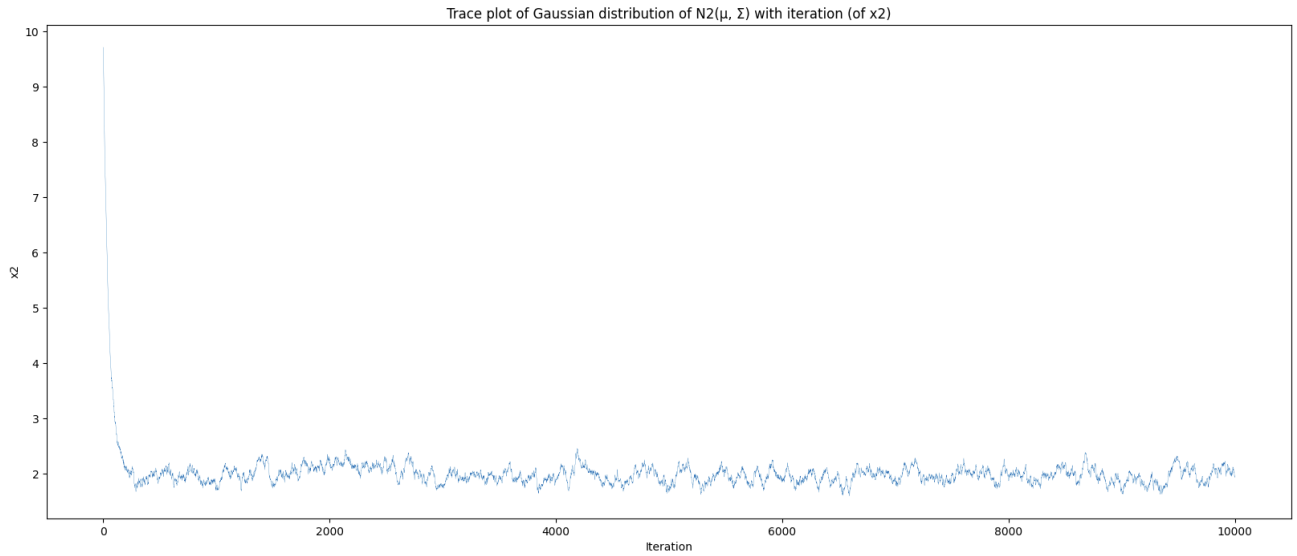


Fig 7. [Trace plot of x2 w.r.t iterations, for a=0.99]

In above trace plots [Fig 6,7], although the initial values are very poor (deviated from the original value), the algorithm converges very quickly to the appropriate values in further iterations. Thus initial few iterations can be also discarded w.r.t the input given.

So from the above graph, burn in of 200 iterations can be used to give the accurate result of the traceplot.
So after eliminating first 200 samples: [Manually set 200 iteration on the basis of traceplot]

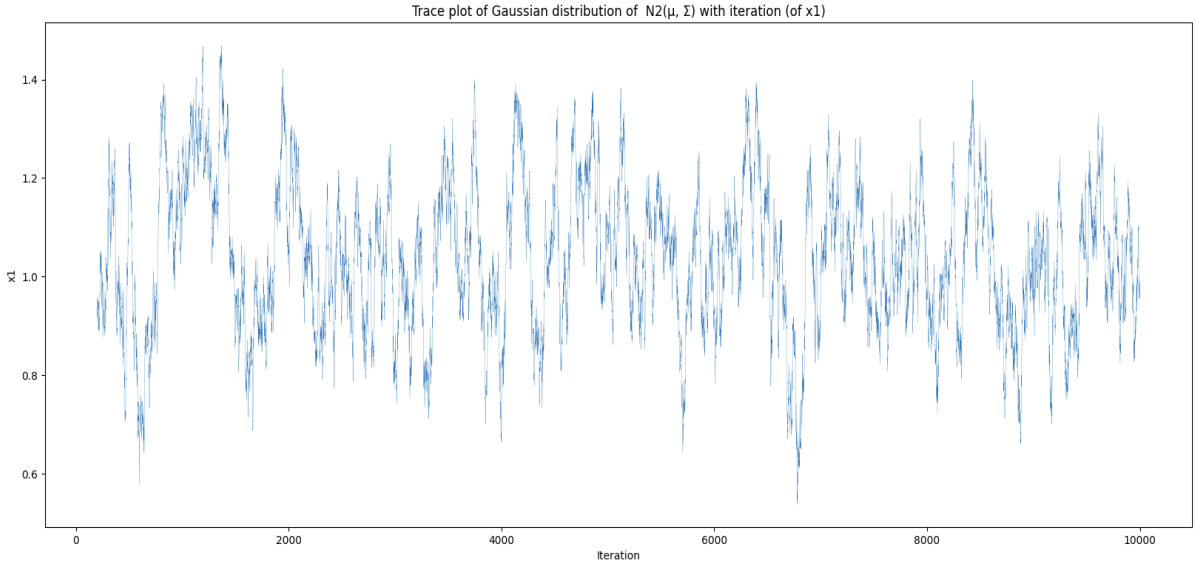


Fig: 8 x1 vs Iterations [a=0.99 and eliminating first 200 samples]

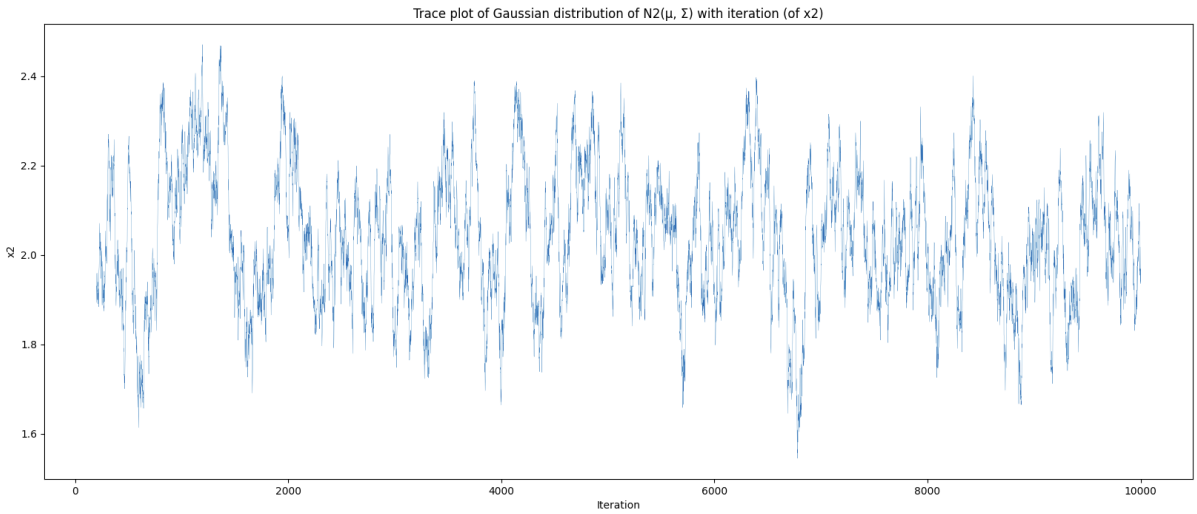


Fig: 9 x2 vs Iterations [a=0.99 and eliminating first 200 samples]