Gibbs Sampling for 2D Joint Distributions

Apurva Nakade

2025-03-05

Table of contents

Gibbs Sampling

Gibbs Sampling is a MCMC algorithm that is used to sample from a joint distribution using conditional distributions. For now, we'll focus on sampling in 2D, but the algorithm generalizes to higher dimensions.

The Gibbs sampling algorithm for sampling from a joint distribution $f_{X,Y}(X, Y)$ is as follows:

1. Start with some initial values X_0 and Y_0.
2. For $i = 1, 2, \dots, N$
    1. Sample $X_i \sim f_{X|Y}(X|Y_{i-1})$.
    2. Sample $Y_i \sim f_{Y|X}(Y|X_i)$.
3. Return the (X_0, Y_0), (X_1, Y_1), ..., (X_N, Y_N).

The algorithm generates a sample path of length N of the Markov Chain as described in Section . If needed, we can discard the initial samples to ensure that the Markov Chain has converged to the stationary distribution.

Joint Exponential Distribution

Let X and Y be two random variables with the truncated exponential distribution:

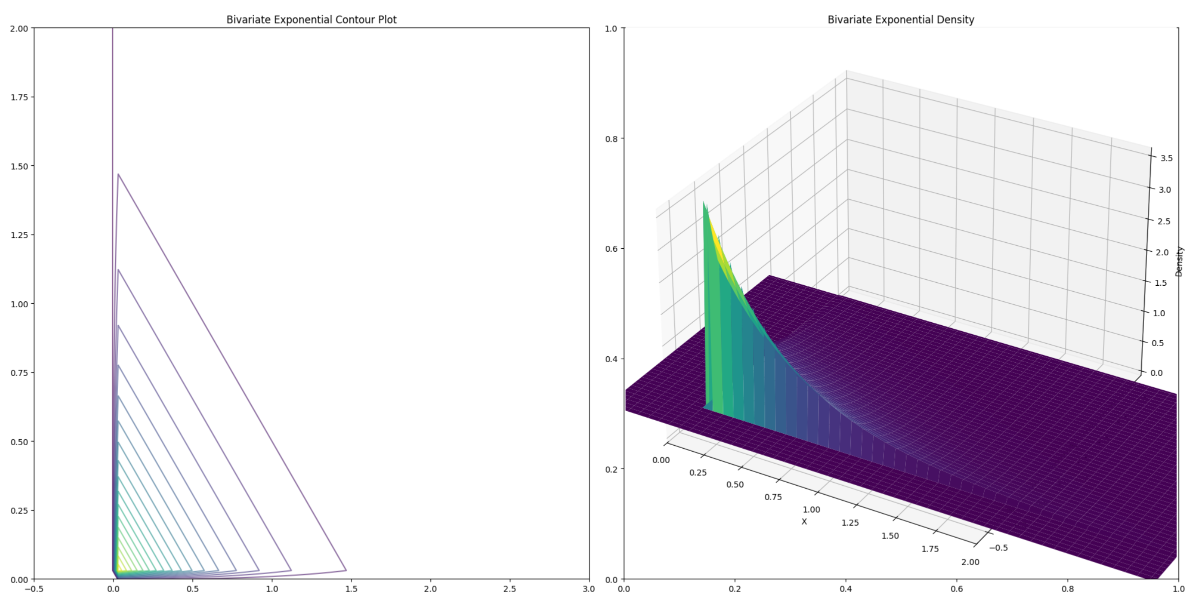
$$f_{X,Y}(x,y) = ce^{-\lambda xy} \text{ for } 0 \leq x \leq D_1, 0 \leq y \leq D_2,$$

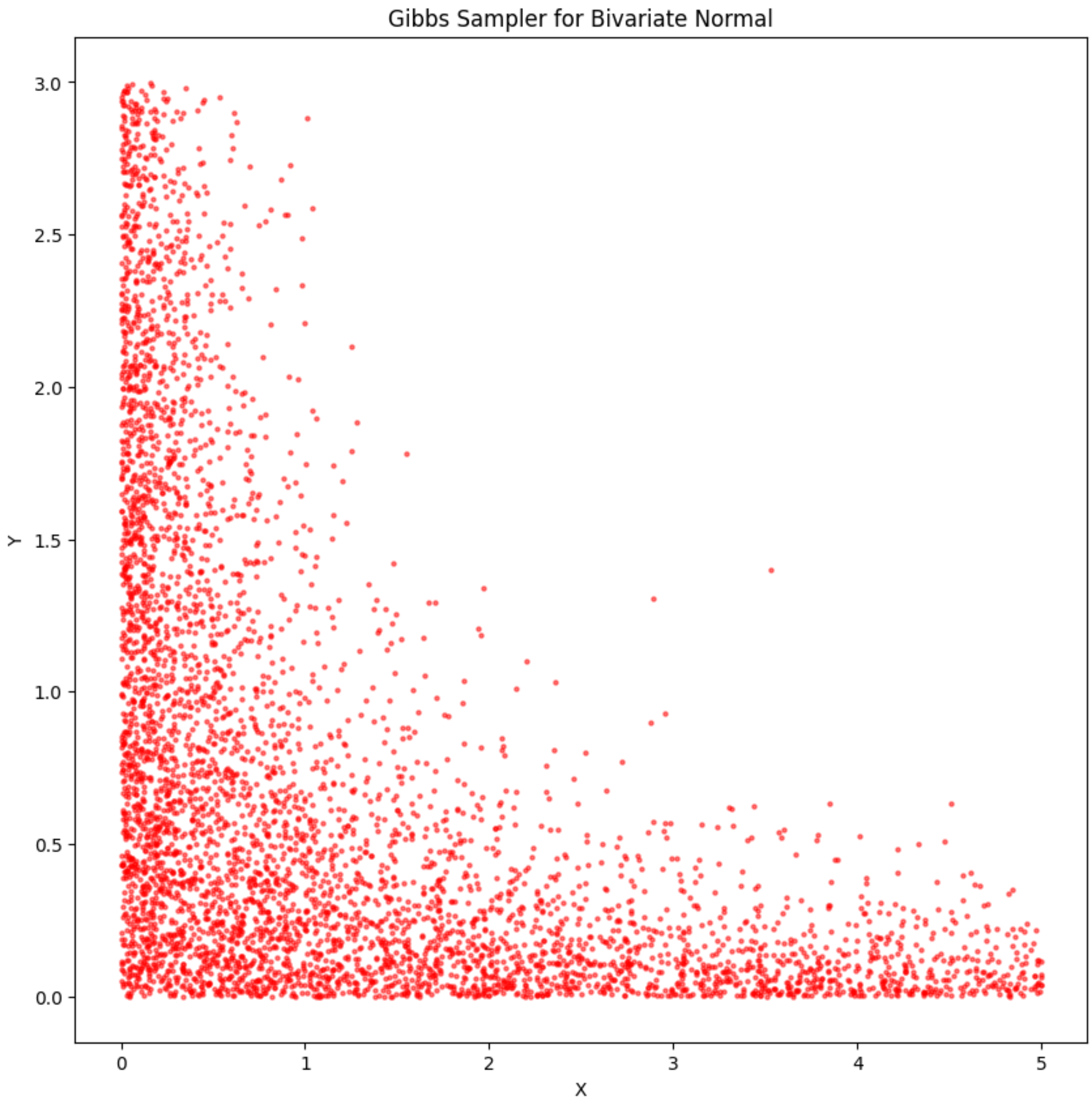
where c is the normalization constant. We want to sample from the joint distribution $f_{X,Y}$. One can show that the conditionals are,

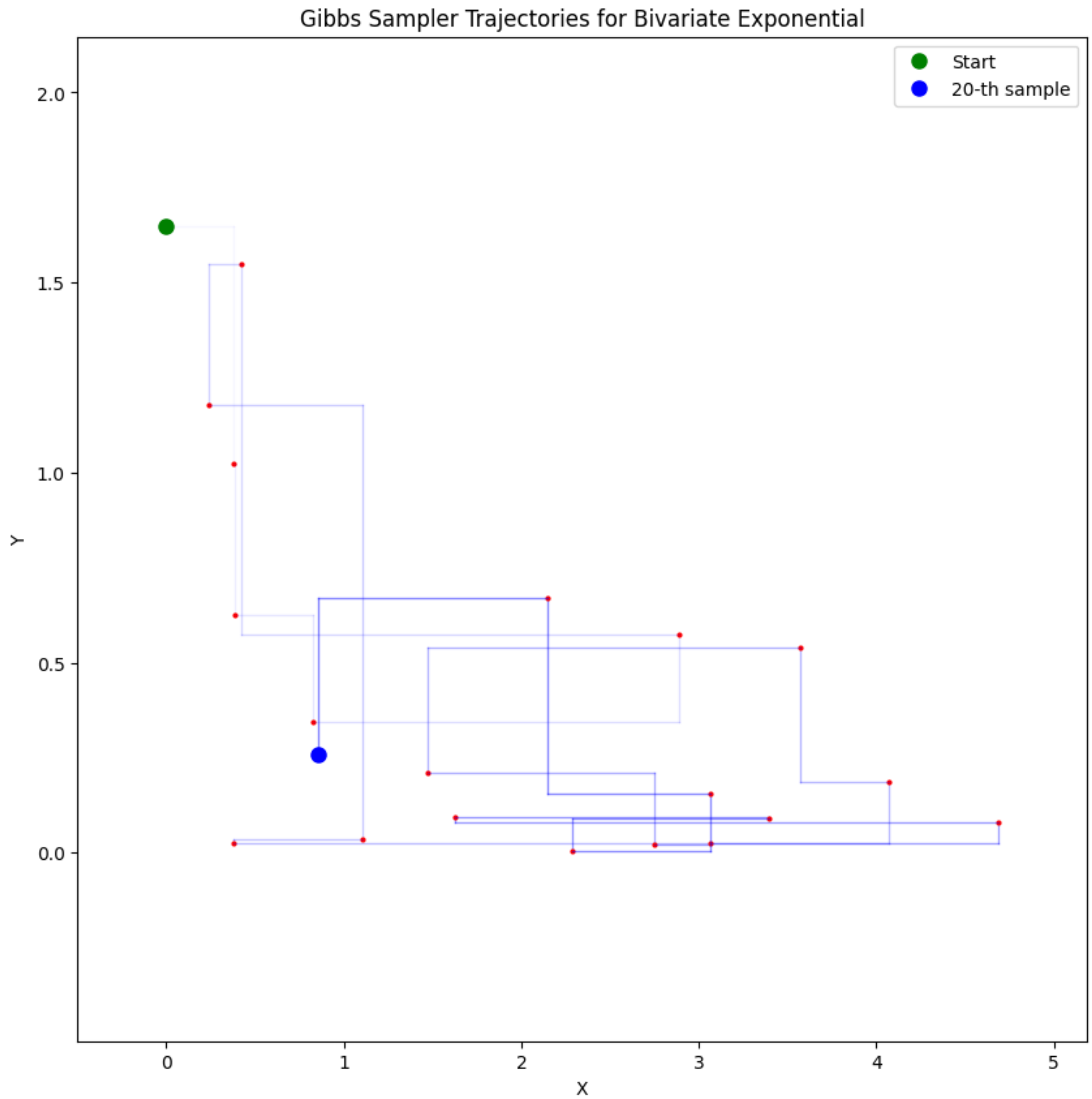
$$\begin{aligned} f_{X|Y}(x|y_0) &= c_{y_0} e^{-\lambda x y_0} \text{ for } 0 \leq x \leq D_1 \\ f_{Y|X}(y|x_0) &= c_{x_0} e^{-\lambda y x_0} \text{ for } 0 \leq y \leq D_2, \end{aligned}$$

where c_{y_0} and c_{x_0} are the normalization constants. We can easily sample from the two conditionals $f_{X|Y}$ and $f_{Y|X}$ using the inverse method function (even for truncated distributions). The Gibbs algorithm becomes

1. Start with some initial values X_0 and Y_0.
2. For $i = 1, 2, \ldots, N$
    1. Sample $X_i \sim c_{Y_{i-1}} e^{-\lambda X_i Y_{i-1}}$ using the inverse method.
    2. Sample $Y_i \sim c_{X_i} e^{-\lambda Y_i X_i}$ using the inverse method.
3. Return the (X_0, Y_0), (X_1, Y_1), ..., (X_N, Y_N).




Gibbs Sampler for Bivariate Normal
Y
X
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5



Bivariate Normal Distribution

Suppose (X, Y) has a bivariate normal distribution with mean $(0, 0)$ and covariance matrix $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. We want to sample from the joint distribution $f_{X,Y}$. One can show that

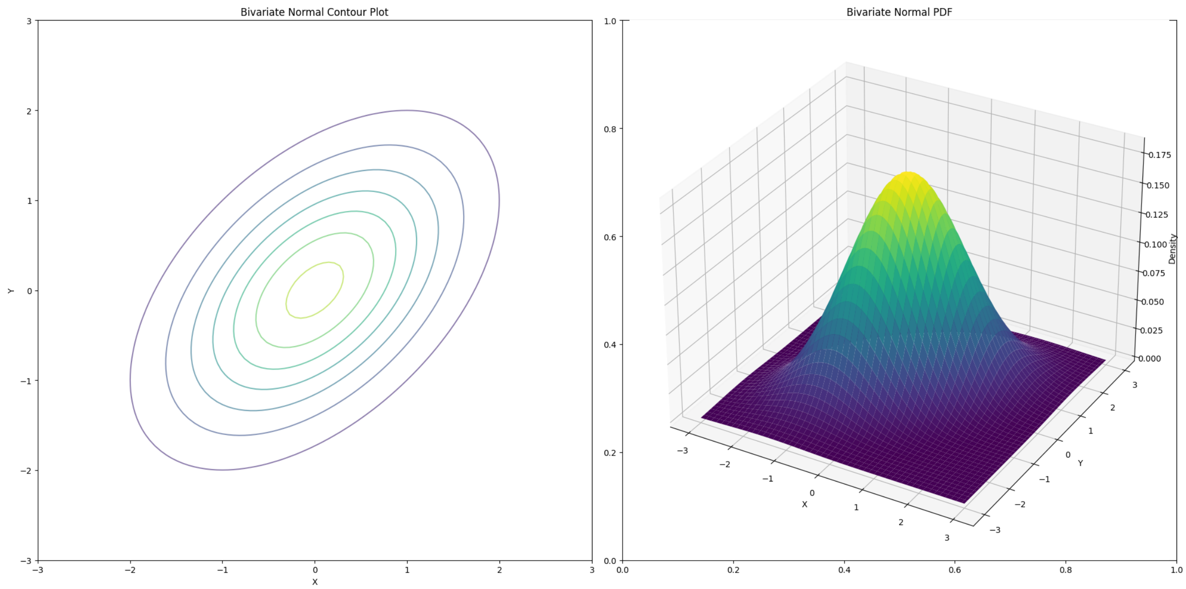
$$(X|Y = y_0) \sim \mathcal{N}(\rho y_0, 1 - \rho^2)$$
$$(Y|X = x_0) \sim \mathcal{N}(\rho x_0, 1 - \rho^2).$$

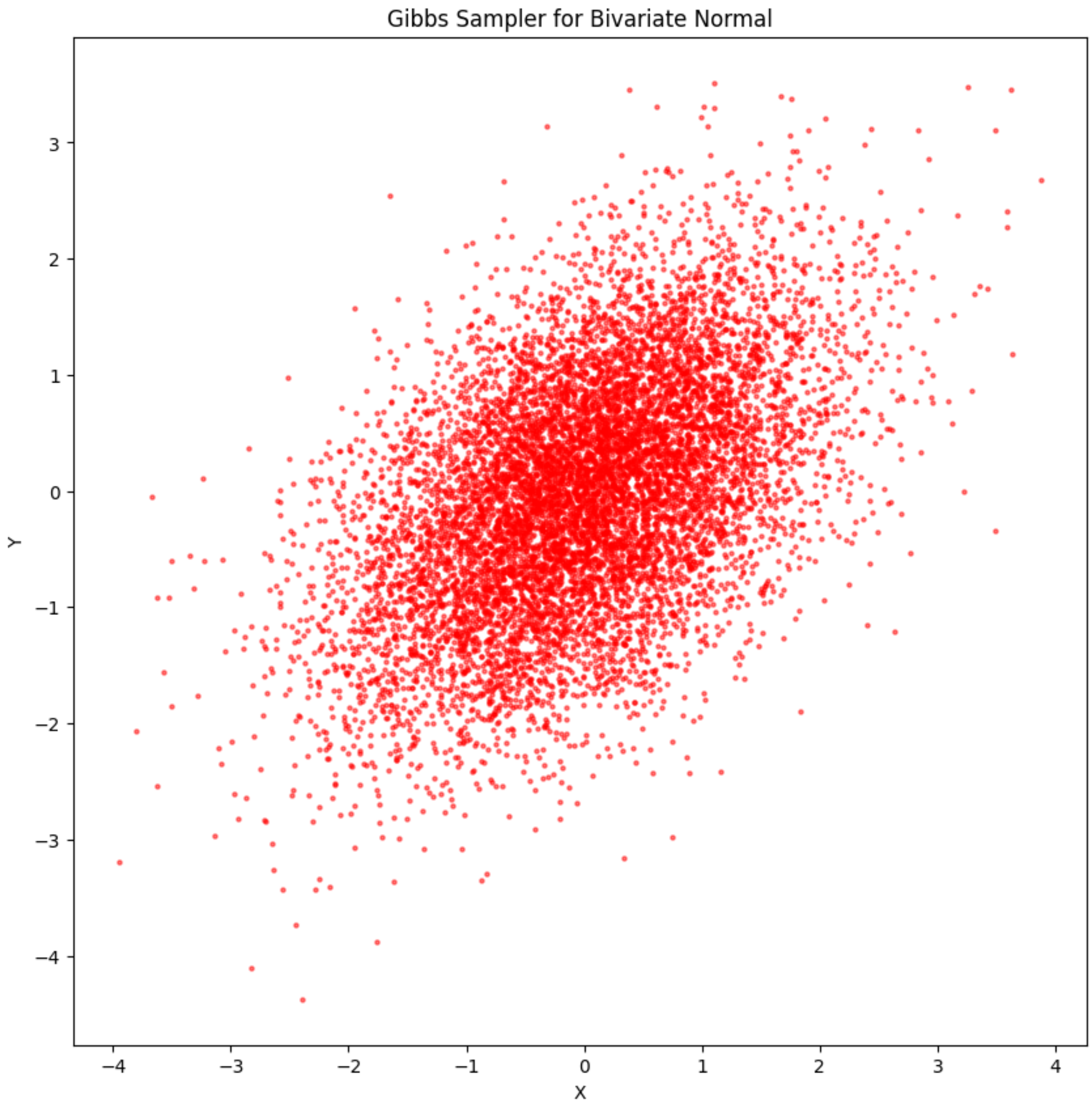
We can sample from the two conditionals $f_{X|Y}$ and $f_{Y|X}$ using the Box-Muller method. The Gibbs algorithm becomes

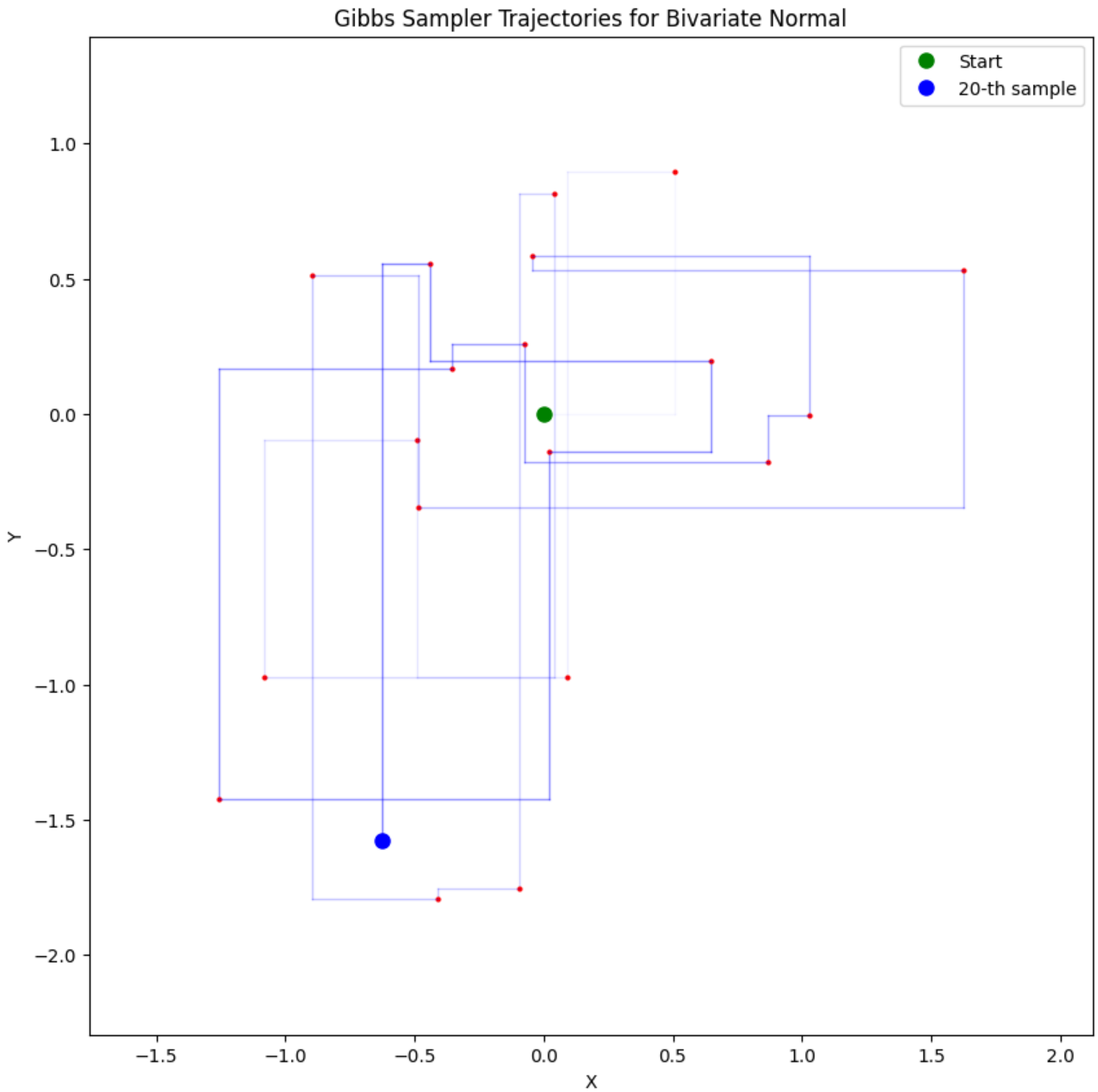
1. Start with some initial values X_0 and Y_0.
2. For $i = 1, 2, \ldots, N$

    1. Sample $X_i \sim \mathcal{N}(\rho Y_{i-1}, 1 - \rho^2)$.
    2. Sample $Y_i \sim \mathcal{N}(\rho X_i, 1 - \rho^2)$.

3. Return the (X_0, Y_0), (X_1, Y_1), ..., (X_N, Y_N).




Gibbs Sampler for Bivariate Normal
Y
X
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3
4



Markov Chain

The Gibbs sampling algorithm generates a Markov Chain whose state space is the product space of the state spaces of the individual variables $\Omega = \Omega_X \times \Omega_Y$. In the above examples, the state space is $[0, D_1] \times [0, D_2]$ for the exponential distribution and $\mathbb{R}^2$ for the bivariate normal distribution. The transition matrix of the Markov Chain in discrete case is given by

$$P\left(\begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} x' \\ y' \end{bmatrix}\right) = \mathbb{P}(X_{i+1} = x' | Y_i = y)\mathbb{P}(Y_{i+1} = y' | X_i = x').$$

In the continuous case, the transition *kernel* is given by

$$K\left(\begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} x' \\ y' \end{bmatrix}\right) = f_{X|Y}(x|y)f_{Y|X}(y|x').$$

Theorem 0.1. *The Gibbs sampling algorithm generates a Markov Chain with the transition matrix P as described above. The joint distribution $f_{X,Y}$ is a stationary distribution of the Markov Chain. Hence, if the Markov Chain converges to the stationary distribution, the samples generated by the Gibbs algorithm will be distributed according to $f_{X,Y}$.*

We'll provide a proof of the above theorem in the next section.