Exercise Walkthrough: Inverse Transform Sampling for a 2D Gaussian

Justin Lanfermann

25. June 2025

Abstract

This document provides a detailed, step-by-step walkthrough for an exercise on sampling from a multivariate Gaussian distribution. It covers the derivation of a method to sample from a standard one-dimensional Gaussian, and extends this to the two-dimensional case using an affine transformation based on the Cholesky decomposition. All steps are justified with references to concepts from the "Discrete Probability Theory" script.

1 Overview of the Problem

The goal of this exercise is to generate samples from a specific 2-dimensional Gaussian distribution:

$$X \sim \mathcal{N}\left(\mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \right)$$

The method we are asked to use is **inverse transform sampling** [1]. This technique, formalized in Proposition 2.23 of the script, provides a general way to turn samples from a uniform distribution $U(0,1)$ into samples from any other distribution, provided we can compute the inverse of its cumulative distribution function (CDF), also known as the quantile function.

Our approach will be:

1. First, figure out how to generate samples from the simplest Gaussian, the standard normal distribution $\mathcal{N}(0,1)$.

2. Then, use an **affine transformation** [2] to convert these standard normal samples into samples from our target 2D Gaussian distribution. This step will involve the **Cholesky decomposition** [3] of the covariance matrix Σ.

2 Part (i): Sampling from a 1D Standard Normal

2.1 The Challenge with Direct Inversion

The inverse transform sampling method (Remark 2.24) states that if we have a random variable $U \sim U(0,1)$, we can generate a sample x from a distribution with CDF F_X by computing $x = F_X^{-1}(u)$.

For a standard normal distribution $Z \sim \mathcal{N}(0,1)$, the CDF is often denoted by $\Phi(z)$:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Unfortunately, this integral does not have a closed-form elementary solution, and neither does its inverse, $\Phi^{-1}(u)$ (the quantile function, also called the probit function). Therefore, we cannot directly apply the inverse transform sampling formula in a simple analytical way.

2.2 A Practical Solution: The Box-Muller Transform

To overcome this, a clever technique called the **Box-Muller transform** [4] is used. It's a beautiful piece of statistics that allows us to generate two independent standard normal samples from two independent uniform samples.

Derivation and Transformation: Let U_1, U_2 be two independent random variables drawn from $U(0, 1)$. The Box-Muller transformation defines two new random variables, Z_1 and Z_2, as follows:

$$Z_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \tag{1}$$
$$Z_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2) \tag{2}$$

It can be proven that Z_1 and Z_2 are independent random variables, and both follow a standard normal distribution, i.e., $Z_1 \sim \mathcal{N}(0, 1)$ and $Z_2 \sim \mathcal{N}(0, 1)$.

This provides the required transformation: to get one standard normal sample, we can generate two uniform samples and apply, for instance, the formula for Z_1. To be efficient, we typically generate pairs of standard normal samples at a time.

3 Part (ii): Sampling from $\mathcal{N}(\mu, \Sigma)$

Now that we can generate samples from a standard normal distribution, we need to transform them to fit our target distribution $\mathcal{N}(\mu, \Sigma)$.

3.1 Using the Affine Transformation Property of Gaussians

Lemma 2.53 (ii) in the script describes a crucial property of Gaussian distributions: an affine transformation of a Gaussian random vector is also a Gaussian random vector. Specifically, if $Z \sim \mathcal{N}(\mu_Z, \Sigma_Z)$ is a d-dimensional Gaussian, then for a matrix $\mathbf{A} \in \mathbb{R}^{k \times d}$ and a vector $\mathbf{b} \in \mathbb{R}^k$, the random vector $Y = \mathbf{A}Z + \mathbf{b}$ is distributed as:

$$Y \sim \mathcal{N}(\mathbf{A}\mu_Z + \mathbf{b}, \mathbf{A}\Sigma_Z\mathbf{A}^\top)$$

Our goal is to find a matrix $\mathbf{A}$ and a vector $\mathbf{b}$ that will transform standard normal samples into our target samples.

Let's start with a 2-dimensional standard normal vector $\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$, where $Z_1, Z_2 \sim \mathcal{N}(0, 1)$ are independent. The distribution of $\mathbf{Z}$ is:

$$\mathbf{Z} \sim \mathcal{N}\left(\mu_Z = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma_Z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}\right)$$

We want to find $\mathbf{A}$ and $\mathbf{b}$ such that $\mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{b}$ follows our target distribution $\mathcal{N}(\mu, \Sigma)$. Using the transformation rule:

$$\mathbf{A}\mathbf{Z} + \mathbf{b} \sim \mathcal{N}(\mathbf{A}\mathbf{0} + \mathbf{b}, \mathbf{A}\mathbf{I}\mathbf{A}^\top) = \mathcal{N}(\mathbf{b}, \mathbf{A}\mathbf{A}^\top)$$

By comparing this with our target distribution $\mathcal{N}(\mu, \Sigma)$, we can see that we need to find $\mathbf{A}$ and $\mathbf{b}$ such that:

1. $\mathbf{b} = \mu = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$

2. $\mathbf{A}\mathbf{A}^\top = \Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

3.2 Finding the Transformation Matrix via Cholesky Decomposition

The second condition, $\mathbf{A}\mathbf{A}^\top = \Sigma$, is a classic problem that can be solved using the Cholesky decomposition. The Cholesky decomposition of a positive-definite matrix Σ finds a unique lower-triangular matrix $\mathbf{L}$ such that $\mathbf{L}\mathbf{L}^\top = \Sigma$. We can simply choose our transformation matrix $\mathbf{A}$ to be this matrix $\mathbf{L}$.

Let's compute the Cholesky decomposition for our covariance matrix $\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$. We are looking for a matrix $\mathbf{L}$ of the form:

$$\mathbf{L} = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix}$$

such that $\mathbf{L}\mathbf{L}^\top = \Sigma$.

$$\mathbf{L}\mathbf{L}^\top = \begin{pmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{pmatrix} \begin{pmatrix} l_{11} & l_{21} \\ 0 & l_{22} \end{pmatrix} = \begin{pmatrix} l_{11}^2 & l_{11}l_{21} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 \end{pmatrix}$$

By equating this with Σ, we get a system of equations:

- $l_{11}^2 = 2 \implies l_{11} = \sqrt{2}$

- $l_{11}l_{21} = 1 \implies \sqrt{2} \cdot l_{21} = 1 \implies l_{21} = \frac{1}{\sqrt{2}}$

- $l_{21}^2 + l_{22}^2 = 2 \implies \left(\frac{1}{\sqrt{2}}\right)^2 + l_{22}^2 = 2 \implies \frac{1}{2} + l_{22}^2 = 2 \implies l_{22}^2 = \frac{3}{2} \implies l_{22} = \sqrt{\frac{3}{2}}$

So, our transformation matrix is:

$$\mathbf{A} = \mathbf{L} = \begin{pmatrix} \sqrt{2} & 0 \\ \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} \end{pmatrix}$$

4 Summary of the Full Algorithm

We now have all the components to generate a sample $\mathbf{X}$ from $\mathcal{N}(\mu, \Sigma)$.

1. **Generate Uniform Samples:** Draw two independent samples, U_1 and U_2, from a $U(0,1)$ distribution.

2. **Generate Standard Normal Samples:** Use the Box-Muller transform to convert U_1, U_2 into a vector of two independent standard normal samples $\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$.

$$Z_1 = \sqrt{-2\ln(U_1)}\cos(2\pi U_2) \quad \text{and} \quad Z_2 = \sqrt{-2\ln(U_1)}\sin(2\pi U_2)$$

3. **Apply Affine Transformation:** Use the computed matrix $\mathbf{L}$ and the given mean vector μ to transform $\mathbf{Z}$ into the final sample $\mathbf{X}$.

$$\mathbf{X} = \mathbf{L}\mathbf{Z} + \mu = \begin{pmatrix} \sqrt{2} & 0 \\ \frac{1}{\sqrt{2}} & \sqrt{\frac{3}{2}} \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Repeating these steps will generate a set of samples that are distributed according to the desired 2-dimensional Gaussian distribution.

Further Explanations

[1] Inverse Transform Sampling

This is a fundamental method for generating random numbers from a given probability distribution, provided its cumulative distribution function (CDF) can be inverted. As stated in Proposition 2.23 and Remark 2.24 of the script, if a random variable X has a continuous and strictly increasing CDF, F_X, then the random variable $U = F_X(X)$ is uniformly distributed on $[0, 1]$. By inverting this relationship, we get $X = F_X^{-1}(U)$. This gives a practical way to sample X: first, draw a sample u from $U(0, 1)$, then compute $x = F_X^{-1}(u)$. The set of such x's will be distributed according to F_X.

[2] Affine Transformation of Gaussians

This property, highlighted in Lemma 2.53 (ii), is a cornerstone of working with multivariate Gaussian distributions. It states that if you take a Gaussian random vector and apply a linear transformation (matrix multiplication) and a shift (vector addition), the result is another Gaussian random vector. This makes Gaussians incredibly tractable, as many complex operations reduce to simple linear algebra on their mean vectors and covariance matrices.

[3] Cholesky Decomposition

The Cholesky decomposition is a numerical analysis technique for decomposing a symmetric, positive-definite matrix into the product of a lower-triangular matrix and its transpose ($\Sigma = \mathbf{L}\mathbf{L}^\top$). It is essentially a "matrix square root". In statistics, it is widely used to find a linear transformation that can turn uncorrelated random variables into correlated ones with a specific covariance structure, which is exactly what was needed in this exercise.

[4] The Box-Muller Transform

The Box-Muller transform is a pseudo-random number sampling method for generating pairs of independent, standard normally distributed random numbers from a single pair of independent, uniformly distributed random numbers. The method is a direct consequence of a change of variables in a two-dimensional Cartesian system to a polar coordinate system. It elegantly bypasses the problem that the standard normal CDF cannot be inverted in a simple closed form.